COMPLETE BIOSYNTHETIC GENE SET FOR SYNTHESIS OF POLYKETIDE ANTIBIOTICS, INCLUDING THE ALBICIDIN FAMILY, RESISTANCE GENES, AND USES THEREOF

This application claims the benefit of U.S. Provisional patent application with Serial No. 60/419,463, filed October 18, 2002 the disclosure of which is hereby incorporated by reference in its entirety, including all nucleic acid sequences, amino acid sequences, chemical formulae, tables and figures.

TECHNICAL FIELD

[0001] The invention is in the field of genetic engineering, and in particular the isolation and expression of the biosynthetic genes that produce a family of antibiotics known generically as albicidins.

BACKGROUND OF THE INVENTION

[0002] U.S. Patent No. 4,525,354 to Birch and Patil described a "non-peptide" antibiotic of M.W. "about 842" called "albicidin." Albicidin is described as produced by culturing chlorosis-inducing strains of *Xanthomonas albilineans* isolated from diseased sugarcane, and mutants thereof. The antibiotic was isolated from the culture medium by adsorption on resin and was purified by gel filtration and High Performance Liquid Chromatography (HPLC). The chemical structure of this antibiotic was not determined and remained unknown, although the Birch and Patil patent disclosed spectral data for a fraction having antibiotic activity and the presence of approximately 38 carbon atoms and at least one COOH group.

[0003] *Xanthomonas albilineans* is a systemic, xylem-invading pathogen that causes leaf scald disease of sugarcane (interspecific hybrids of *Saccharum* species) (Ricaud and Ryan, 1989; Rott and Davis, 2000). Leaf scald symptoms include chlorosis, necrosis, rapid wilting, and plant death. Chlorosis-inducing strains of the pathogen produce several toxic compounds. The major toxic component, named albicidin, inhibits chloroplast DNA replication, resulting in blocked chloroplast differentiation and chlorotic leaf streaks that are characteristic of the plant disease (Birch and Patil, 1983, 1985b, 1987a and 1987b). Several studies established that albicidin plays a key role in pathogenesis and especially in the development of disease symptoms (Wall and Birch, 1997; Zhang and Birch, 1997; Zhang *et al.*, 1999; Birch, 2001).

[0004] The prior art indicates that albicidin inhibits prokaryotic DNA replication and is bactericidal to a range of gram-positive and gram-negative bacteria (Birch and Patil, 1985a). Albicidin is therefore of interest as a potential clinical antibiotic (Birch and Patil, 1985a). However, low yield of toxin production in *X. albilineans* has slowed down studies into the chemical structure of albicidin and its therapeutic application (Zhang *et al.*, 1998). The chemical structure of this albicidin remains unknown, however this albicidin has been partially

characterized as a non-peptide antibiotic with a molecular weight of about 842 that contains approximately 38 carbon atoms with three or four aromatic rings, at least one COOH group, two OCH3 groups, a trisubstituted double bond and a CN linkage (Birch and Patil, 1985a; Huang *et al.*, 2001).

[0005] Molecular cloning and characterization of the genes governing the biosynthesis of albicidin is of considerable interest because such information provides approaches to engineer overproduction of albicidin, to characterize its chemical structure, to allow therapeutic applications and to clarify the relationship between toxin production and the ability to colonize sugarcane. Two similar mutagenesis and complementation studies have been conducted to identify the genetic basis of albicidin production in *X. albilineans* strains isolated in two different geographical locations, Australia and Florida.

[0006] One study of *X. albilineans* strain LS155 from Australia revealed that genes for albicidin biosynthesis and resistance span at least 69kb (Wall and Birch, 1997). Subsequently, three genes required for albicidin biosynthesis were identified, cloned and sequenced from two Australian strains of *X. albilineans* (LS155 and Xa13): *xabA*, *xabB* and *xabC* (Huang et al., 2001; Huang et al. 2000a, 2000b). The *xabB* gene encodes a large protein with a predicted size of 525.6 kDa, with a modular architecture indicative of a multi functional polyketide synthase (PKS) linked to a nonribosomal peptide synthetase (NRPS) (Huang *et al.*, 2001). The *xabC* gene, located immediately downstream from *xabB*, encodes an S-adenosyl-L-methionine (SAM)-dependent *O*-methyltransferase (Huang *et al.*, 2000a). The *xabA* gene, located in another region of the genome, encodes a phosphopantetheinyl transferase required for post-translational activation of PKS and NRPS enzymes (Huang *et al.*, 2000b).

[0007] These first results demonstrated that the albicidin biosynthesis apparatus is a PKS and/or NRPS system. Such systems assemble simple acyl-coenzyme A or amino acid monomers to produce polyketides and/or nonribosomal peptides (Marahiel *et al.*, 1997; Cane, 1997; Cane and Walsh, 1999). These metabolites form very large classes of natural products that include numerous important pharmaceuticals, agrochemicals, and veterinary agents such as antibiotics, immunosuppressants, anti-cholesterolemics, as well as antitumor, antifungal and antiparasitic agents. Genetic studies of prokaryotic PKS and NRPS produced detailed information regarding the function and the organization of genes responsible for the biosynthesis of polyketides and nonribosomal peptides. Such knowledge, in turn, made it possible to produce combinations of PKS and NRPS genes from different microorganisms in order to produce novel antibiotics (McDaniel *et al.*, 1999; Rodriguez and McDaniel, 2001; Pfeifer *et al.*, 2001). Investigating the complete albicidin biosynthesis apparatus is therefore of great interest because such results may contribute to the knowledge as to how PKS and NRPS interact and how they might be manipulated to engineer novel molecules.

[0008] A second study with *X. albilineans* strain Xa23R1 from Florida revealed that at least two gene clusters, one spanning more than 48 kb, are involved in albicidin production (Rott *et al.*, 1996). This conclusion was based on the following data: (I) fifty Xa23R1 mutants defective in albicidin production were isolated; (ii) a Xa23R1 genomic library of 845 clones, designated pALB1 to pALB845, was constructed; (iii) two overlapping DNA inserts of approximately 47 kb and 41 kb, from clones pALB540 and pALB571 respectively, complemented forty-five mutants and were supposed to contain a major gene cluster involved in albicidin production; (iv) a DNA insert of approximately 36 kb, from clone pALB639, complemented four of the five remaining mutants not complemented by pALB540 and pALB571, and was supposed to contain a second region involved in albicidin production; and (v) the remaining mutant, AM37, which was not complemented by any of the three cosmids pALB540, pALB571 and pALB639, was supposed to be mutated in a third region of the genome involved in albicidin production.

[0009] The DNA sequences of all of the genes required to produce the albicidin family of polyketide antibiotics, the expressed protein amino acid sequences of all of the genes, and the deduced structure of Albicidin have not been previously reported, although fragmentary sequences that include three of the biosynthetic genes have been reported. Identification of one albicidin gene, *xabC*, as a methyltransferase gene involved in albicidin biosynthesis is reported by Huang, G., Zhang, L. & Birch, R.G. (2000a, Gene **255**, 327-333) and claimed as biologically active in producing a polyketide antibiotic in PCT WO 02/24736 A1. Identification of a second albicidin gene, *xabA*, as a phosphopantetheinyl transferase gene is reported by Huang, G., Zhang, L. and Birch, R.G. (2000b) Gene **258**, 193-199 and claimed as biologically active in producing a polyketide antibiotic in PCT WO 02/24736 A1. Huang, G., Zhang, L. & Birch, R.G. (2001) Microbiology **147**, 631-642, report a DNA sequence of *xabB* (GenBank accession # AF239749), a multi functional polyketide-peptide synthetase that may be essential for albicidin biosynthesis in *Xanthomonas albilineans.* This *xabB* gene is reported as full length by Birch in PCT WO 02/24736 A1 (their seq. ID #1) and claimed by Birch in PCT WO 02/24736 A1 as a biologically active polyketide synthase of 4,801 amino acids in length, enabling production of albicidin. However, the DNA sequence reported by Huang *et al.* (2001) in GenBank AF239749 and by Birch in PCT WO 02/24736 A1 (their seq. ID #1) appears to be incomplete and missing 6,234 bp of DNA sequence encoding 2,078 amino acids. The subject invention provides the complete DNA sequence of *xabB* (*albI*, our seq. 20) as 20,637 bp, encoding a biologically active polyketide synthase of 6,879 amino acids of in this application (our seq ID #26). Factors affecting biosynthesis by *Xanthomonas albilineans* of albicidins antibiotics and phytotoxins are discussed in J. Appl. Microbiol. **85**, 1023-1028. and Wall, M.K. & Birch, R.G. (1997). Genes for albicidin biosynthesis and resistance span at least 69 kb in the genome of *Xanthomonas albilineans*. Lett. Appl. Microbiol. **24**, 256-260. A gene from *X. albilineans* strain Xa13, designed AlbF, which confers high level albicidin resistance in *Escherichia coli* and which encodes a putative albicidin

efflux pump, was directly submitted to Genbank by Bostock and Birch (Accession No. AF403709).

SUMMARY OF THE INVENTION

[0010] The present invention describes and characterizes the family of antibiotics that is produced by culturing chlorosis-inducing strains of *X. albilineans* and mutants thereof, together with the complete set of twenty biosynthetic genes capable of producing the unique and previously uncharacterized family of antibiotics produced by *X. albilineans* and previously lumped together as "albicidins." The set of twenty biosynthetic genes isolated, purified and cloned from a culture of *X. albilineans* revealed that this set of biosynthetic genes is capable of synthesizing products exhibiting a high level of variation among the products, indicating that albicidins comprise a family of polyketide antibiotics. The albicidins described in the present invention are synthesized by twenty genes, including one polyketide-peptide synthase, one polyketide synthase and two peptide synthases, but the substrates of the polyketide-peptide synthase and of one peptide synthase are not α-amino acids. The biosynthetic enzymes represent a previously undescribed and unique polyketide antibiotic biosynthetic system.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] **Figure 1** is a Physical Map and genetic organization of the DNA Region containing the major gene cluster XALB1 involved in the biosynthesis of Albicidins.

[0012] **Figure 2** is an illustration of the organization of the four PKS modules and the seven NRPS modules identified in cluster XALB1 and comparison with the organization of the prior art material XabB.

[0013] **Figure 3** shows the conserved sequence motifs in O-methyltransferases and C-methyltransferases involved in antibiotic biosynthesis in bacteria and in AlbII.

[0014] **Figure 4** shows the conserved sequence motifs in O-methyltransferases and in different tcmP-like hypothetical proteins and AlbVI.

[0015] **Figure 5** is an illustration of the alignment of the primary sequences between the conserved motifs A4 and A5 of Alb NPRSs and PKS-4 in *Xanthomonas albilineans* with the corresponding sequences of GrsA (Phe) accession number: P14687 and Blm NRPS-2 (β-Ala) accession number AF210249.

[0016] **Figure 6** shows Rho-independent transcription terminators identified in the intergenic regions of XALB1 and XALB3 clusters.

[0017] **Figure 7A** shows sequences identified as a putative bidirectional promoter between albX and albXVII in XALB1 for transcriptional control of operons 3 and 4.

[0018] Figure 7B shows sequences identified as a putative unidirectional promoter upstream from *albXIX* for transcriptional control of operon 5 if *albXVIII* is not expressed.

[0019] Figure 8 is a physical map and genetic organization of the DNA region containing the gene clusters XALB2 and XALB3 involved in albicidin production.

[0020] Figure 9A is linear model 1 leading to the biosynthesis of only one polyketide-polypeptide albicidin backbone.

[0021] Figure 9B is linear model 2 leading to the biosynthesis of four different polyketide-polypeptide backbone.

[0022] Figure 10A is an alignment of the conserved motifs in AT domains from RifA-1, -2, -3, RifB-1, RifE-1 (Rifamycin PKSs, August *et al.*, 1998) and BlmVIII (Bleomycin PKS; Du *et al.*, 2000).

[0023] Figure 10B is a comparison of AlbXIII, FenF (a malonyl-CoA transacylase located upstream from *mycA*, Duitman *et al.*, 1999) and LipA (a lipase; Valdez *et al.*, 1999).

[0024] Figure 11A is a proposed model for biosynthesis of albicidin, including putative substrates of PKS and NRPS modules.

[0025] Figure 11B shows the proposed compositions and structures of albicidins.

[0026] Figure 12 illustrates subcloning of operons 3 and 4 (from pALB540), XALB2 (from pAC389.1) and XALB3 (from pEV639) into a single plasmid, pOp3-4/XALB2-3. A *Bam*HI–*Pst*I fragment from pALB540, corresponding to a portion of operon 4, was subcloned into pBCKS(+), yielding pBC/Op4D (step 1). A *Xho*I site was introduced into this vector immediately upstream from the *Bfr*I site by directed mutagenesis, yielding pBC/Op4DXhoI (step 2). The *Eco*RI fragment from pAC389.1 (XALB2) was then subcloned into pBC/Op4DXhoI, yielding pBC/Op4D/XALB2 (step 3). A *Bfr*I fragment from pALB540 containing complete operon 3 and the beginning of operon 4 was subcloned into pBC/Op4D/XALB2, yielding pBC/Op3-4/XALB2 (step 4). The *Sal*I fragment from pEV639 (XALB3) was subcloned into pBKS, yielding pBKS/XALB3 (step 5). The *Sal*I site located on the *Kpn*I side of the polylinker was then destroyed and substituted by a *XhoI* restriction site, yielding pBKS/XALB3XhoI (step 6). Finally, the *Xho*I cassette of pBC/Op3-4/XALB2 was subcloned into the *Sal*I restriction site of pBKS/XALB3XhoI, yielding pBKS/Op3-4/XALB2-3 (step 7). An *Xho*I site was added to the BamHI site of pLAFR3, yielding pLAFR3XhoI (step8). The *Xho*I cassette from pBKS/Op3-4/XALB2-3 was then cloned into pLAFR3XhoI, yielding pOp3-4/XALB2-3 (step 9).

DETAILED DESCRIPTION OF THE INVENTION

[0027] The invention results from the DNA sequencing of the complete major gene cluster XALB1, as well as the noncontiguous fragments XALB2 and XALB3. XALB1 is present in the two overlapping DNA inserts of clones pALB540 and pALB571. Reading frame analysis

and homology analyses allow one to predict the genetic organization of XALB1 and to assign a function to the genes potentially required for albicidin production. Based on the alignment of the different PKS and/or NRPS enzymes encoded by XALB1 we proposed a model for the albicidin backbone biosynthesis. However the invention disclosed herein does not depend upon the accuracy of the proposed model. The invention includes the successful cloning and DNA sequencing of the second region of the genome (XALB2) involved in albicidin production and mutated in mutant AM37.

[0028] The invention includes the characterization of the third region of the genome (XALB3) involved in albicidin production present in clone pALB639. These results allowed the possibility to characterize all enzymes of the albicidin biosynthesis pathway including structural, resistance and regulatory elements and to engineer overproduction of albicidin.

[0029] The subject invention provides:

(a) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, and 25;

(b) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide encoding a polypeptide selected from the group consisting of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47;

(c) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide that is complementary to a polynucleotide selected from the group consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, and 25;

(d) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide that is complementary to a polynucleotide encoding a polypeptide selected from the group consisting of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47; or

(e) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide that is at least 70% homologous to: (1) a polynucleotide selected from the group consisting of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, and 25; (2) a polynucleotide sequence encoding a polypeptide selected from the group consisting of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47; (3) a polynucleotide that is complementary to a polynucleotide encoding a polypeptide selected from the group consisting of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47; (3) a polynucleotide that is complementary to a polynucleotide sequence selected from the group consisting of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, and 25;

(f) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide sequence encoding variant (e.g., a variant polypeptide) of a polypeptide selected

from the group consisting of SEQ ID NOs: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47, wherein said variant has at least on of the biological activities associated with the polypeptides of SEQ ID NOs: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47;

g) isolated, recombinant, and/or purified polynucleotide sequences comprising polynucleotide sequence encoding a fragment of a polypeptide selected from the group consisting of SEQ ID NOs: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47 or a fragment of a variant polypeptide of SEQ ID NOs: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47;

h) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide sequence encoding multimeric construct;

j) a genetic construct comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h);

k) a vector comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h);

l) a host cell comprising a vector a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h);

m) a transformed plant cell comprising a vector comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h);

n) a transformed plant comprising a vector comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h); or;

o) a polynucleotide that hybridizes under low, intermediate or high stringency with a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h).

[0030] "Nucleotide sequence", "polynucleotide" or "nucleic acid" can be used interchangeably and are understood to mean, according to the present invention, either a double-stranded DNA, a single-stranded DNA or products of transcription of the said DNAs (e.g., RNA molecules). It should also be understood that the present invention does not relate to genomic polynucleotide sequences in their natural environment or natural state. The nucleic acid, polynucleotide, or nucleotide sequences of the invention can be isolated, purified (or partially purified), by separation methods including, but not limited to, ion-exchange chromatography, molecular size exclusion chromatography, or by genetic engineering methods such as amplification, subtractive hybridization, cloning, subcloning or chemical synthesis, or combinations of these genetic engineering methods.

[0031] A homologous polynucleotide or polypeptide sequence, for the purposes of the present invention, encompasses a sequence having a percentage identity with the polynucleotide or polypeptide sequences, set forth herein, of between at least (or at least about) 70.00% to 99.99% (inclusive). The aforementioned range of percent identity is to be taken as including, and

providing written description and support for, any fractional percentage, in intervals of 0.01%, between 20.00% and, up to, including 99.99%. These percentages are purely statistical and differences between two nucleic acid sequences can be distributed randomly and over the entire sequence length. For example, homologous sequences can exhibit a percent identity of 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent with the sequences of the instant invention. Typically, the percent identity is calculated with reference to the polynucleotide of a particular SEQ ID NO.; the full-length of a selected polynucleotide, or the native (naturally occurring) polynucleotide. The terms "identical" or percent "identity", in the context of two or more polynucleotide or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using a sequence comparison algorithm or by manual alignment and visual inspection.

[0032] A "complementary" polynucleotide sequence, as used herein, generally refers to a sequence arising from the hydrogen bonding between a particular purine and a particular pyrimidine in double-stranded nucleic acid molecules (DNA-DNA, DNA-RNA, or RNA-RNA). The major specific pairings are guanine with cytosine and adenine with thymine or uracil. A "complementary" polynucleotide sequence may also be referred to as an "antisense" polynucleotide sequence or an "antisense" sequence.

[0033] Sequence homology and sequence identity can also be determined by hybridization studies under high stringency, intermediate stringency, and/or low stringency. Various degrees of stringency of hybridization can be employed. The more severe the conditions, the greater the complementarity that is required for duplex formation. Severity of conditions can be controlled by temperature, probe concentration, probe length, ionic strength, time, and the like. Preferably, hybridization is conducted under low, intermediate, or high stringency conditions by techniques well known in the art, as described, for example, in Keller, G.H., M.M. Manak [1987] DNA Probes, Stockton Press, New York, NY, pp. 169-170.

[0034] It is also well known in the art that restriction enzymes can be used to obtain functional fragments of the subject DNA sequences. For example, Bal31 exonuclease can be conveniently used for time-controlled limited digestion of DNA (commonly referred to as "erase-a-base" procedures). See, for example, Maniatis et al. [1982] Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York; Wei et al. [1983] J. Biol. Chem. 258:13006-13512.

[0035] The present invention further comprises fragments of the polynucleotide sequences of the instant invention. Representative fragments of the polynucleotide sequences according to the invention will be understood to mean any nucleotide fragment having at least 5

successive nucleotides, preferably at least 12 successive nucleotides, and still more preferably at least 15 or at least 20 successive nucleotides of the sequence from which it is derived. The upper limit for such fragments is the total number of nucleotides found in the full-length sequence encoding a particular polypeptide (e.g., a polypeptide selected from the group consisting of SEQ ID NOs: 26-50). The term "successive" can be interchanged with the term "consecutive". In some embodiments, a polynucleotide fragment may be referred to as "a contiguous span of at least X nucleotides, wherein X is any integer value beginning with 5. The upper limit for polynucleotide fragments of the subject invention is the total number of nucleotides found in the full-length sequence of a particular SEQ ID or the total number of nucleotides encoding a particular polypeptide (e.g., a particular SEQ ID NO).

[0036] In some embodiments, the subject invention includes those fragments capable of hybridizing under various conditions of stringency conditions (e.g., high or intermediate or low stringency) with a nucleotide sequence according to the invention; fragments that hybridize with a nucleotide sequence of the subject invention can be, optionally, labeled as set forth below.

[0037] Thus, the subject invention also provides detection probes (e.g., fragments of the disclosed polynucleotide sequences) for hybridization with a target sequence or the amplicon generated from the target sequence. Such a detection probe will comprise a contiguous/consecutive span of at least 8, 9, 10, 11, 12, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides. Labeled probes or primers are labeled with a radioactive compound or with another type of label as set forth above. Alternatively, non-labeled nucleotide sequences may be used directly as probes or primers; however, the sequences are generally labeled with a radioactive element (32P, 35S, 3H, 125I) or with a molecule such as biotin, acetylaminofluorene, digoxigenin, 5-bromo-deoxyuridine, or fluorescein to provide probes that can be used in numerous applications.

[0038] The subject invention also provides for modified nucleotide sequences. Modified nucleic acid sequences will be understood to mean any nucleotide sequence that has been modified, according to techniques well known to persons skilled in the art, and exhibiting modifications in relation to the native, naturally occurring nucleotide sequences.

[0039] The subject invention also provides genetic constructs comprising: a) a polynucleotide sequence encoding a polypeptide sequence selected from the group consisting of SEQ ID NOs: 1-25; b) a polynucleotide sequence having at least about 70% to 99.99% identity to a polynucleotide sequence encoding a polypeptide sequence selected from the group consisting of SEQ ID NO: 26-50, wherein said polynucleotide encodes a polypeptide having at least one of the biological activities of the polypeptides (e.g., a catalytic activity as set forth in Table 4); c) a polynucleotide sequence encoding a biologically active fragment of a polypeptide selected from the group consisting of SEQ ID NO: 26-50, wherein said biologically active fragment has at least

one of the biological activities of the polypeptides (e.g., a catalytic or transport activity as set forth in Table 4); d) a polynucleotide sequence comprising SEQ ID NO: 1, 2, 3, or combinations thereof; e) a polynucleotide sequence encoding variant (e.g., a variant polypeptide) of a polypeptide selected from the group consisting of SEQ ID NOs: 26-48, wherein said variant has at least on of the biological activities associated with the polypeptides (e.g., a catalytic or transport activity as set forth in Table 4); f) a polynucleotide sequence encoding a fragment of a variant polypeptide as set forth in (e); or g) a polynucleotide sequence encoding multimeric construct. Genetic constructs of the subject invention can also contain additional regulatory elements such as promoters and enhancers and, optionally, selectable markers.

[0040] Also within the scope of the subject instant invention are vectors or expression cassettes containing polynucleotides encoding the polypeptides, set forth supra, operably linked to regulatory elements. The vectors and expression cassettes may contain additional transcriptional control sequences as well. The vectors and expression cassettes may further comprise selectable markers. The expression cassette may contain at least one additional gene, operably linked to control elements, to be co-transformed into the organism. Alternatively, the additional gene(s) and control element(s) can be provided on multiple expression cassettes. Such expression cassettes are provided with a plurality of restriction sites for insertion of the sequences of the invention to be under the transcriptional regulation of the regulatory regions. The expression cassette(s) may additionally contain selectable marker genes operably linked to control elements.

[0041] In some embodiments, the expression cassette will include in the 5'-3' direction of transcription, a transcriptional and translational initiation region, a DNA sequence of the invention, and a transcriptional and translational termination region functional in plants. The transcriptional initiation region, the promoter, may be native or analogous, or foreign or heterologous, to the plant host. Additionally, the promoter may be the natural sequence or alternatively a synthetic sequence. By "foreign" is intended that the transcriptional initiation region is not found in the native plant into which the transcriptional initiation region is introduced. As used herein, a chimeric gene comprises a coding sequence operably linked to a transcriptional initiation region that is heterologous to the coding sequence.

[0042] The termination region may be native with the transcriptional initiation region, may be native with the operably linked DNA sequence of interest, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of A. tumefaciens, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau et al. (1991) Mol. Gen. Genet. 262:141-144; Proudfoot (1991) Cell 64:671-674; Sanfacon et al. (1991) Genes Dev. 5:141-149; Mogen et al. (1990) Plant Cell 2:1261-1272; Munroe et al. (1990) Gene 91:151-158; Ballas et al. (1989) Nucleic Acids Res. 17:7891-7903; and Joshi et al. (1987) Nucleic Acid Res. 15:9627-9639.

[0043] Where appropriate, the polynucleotides encoding the polypeptides set forth supra can be optimized for expression in the transformed plant. That is, the genes can be synthesized using plant-preferred codons corresponding to the plant of interest. Methods are available in the art for synthesizing plant-preferred genes. See, for example, U. S. Patent Nos. 5,380,831 and 5,436,391, and Murray et al. (1989) Nucleic Acids Res. 17:477-498, herein incorporated by reference.

[0044] The expression cassettes may additionally contain 5' leader sequences in the expression cassette construct. Such leader sequences can act to enhance translation. Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region), Elroy-Stein et al. (1989) PNAS USA 86:6126-6130; potyvirus leaders, for example, TEV leader (Tobacco Etch Virus), Allison et al. (1986); MDMV Leader (Maize Dwarf Mosaic Virus), Virology 154:9-20; human immunoglobulin heavy-chain binding protein (BiP), Macejak et al. (1991) Nature 353:90-94; untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4), Jobling et al. (1987) Nature 325:622-625; tobacco mosaic virus leader (TMV), Gallie et al. (1989) in Molecular Biology of RNA, ed. Cech (Liss, New York), pp. 237-256; and maize chlorotic mottle virus leader (MCMV), Lommel et al. (1991) Virology 81:382-385. See also, Della-Cioppa et al. (1987) Plant Physiol. 84:965-968. Other methods known to enhance translation can also be utilized.

[0045] Also provided are transformed host cells, transformed plant cells and transgenic plants which contain one or more genetic constructs, vectors, or expression cassettes comprising polynucleotides of the subject invention, or biologically active fragments thereof, operably linked to control elements. As used herein, the term "planta" includes algae and higher plants. Thus, algae, monocots, and dicots may be transformed with genetic constructs of the invention, expression cassettes, or vectors according to the invention. In certain embodiments of the subject invention, the transformed cells or transgenic plants comprise at least one polynucleotide seuquence selected from the group consisting of SEQ ID NOs: 1-25. In certain preferred embodiments, transformed cells or transgenic plants comprise at least one polynucleotide sequence comprising SEQ ID NO: 1, 2, or 3. Optionally, the transformed cells or transgenic plants can comprise at least two or all three polynucleotide sequences selected from the group consisting of SEQ ID NOs: 1, 2, and 3.

[0046] Methods of transforming cells with genetic constructs, vectors, or expression cassettes comprising the novel polynucleotides of the invention are also provided. These methods comprise transforming a plant or plant cell with a polynucleotide according to the subject invention. Plants and plant cells may be transformed by electroporation, Agrobacterium transformation (including vacuum infiltration), engineered plant virus replicons, electrophoresis, microinjection, micro-projectile bombardment, vacuum infiltration of Agrobacterium, micro-

LASER beam-induced perforation of cell wall, or simply by incubation with or without polyethylene glycol (PEG). Plants transformed with a genetic construct of the invention may be produced by standard techniques known in the art for the genetic manipulation of plants. DNA can be transformed into plant cells using any suitable technology, such as a disarmed Ti-plasmid vector carried by Agrobacterium exploiting its natural gene transferability. Agrobacterium transformation is used by those skilled in the art to transform algae and dicotyledonous species. Substantial progress has been made towards the routine production of stable, fertile transgenic plants in almost all economically relevant monocot plants. In particular, Agrobacterium mediated transformation has now emerged as a highly efficient transformation method in monocots. Microprojectile bombardment, electroporation, and direct DNA uptake are preferred where Agrobacterium is inefficient or ineffective. Alternatively, a combination of different techniques may be employed to enhance the efficiency of the transformation process, e.g., bombardment with Agrobacterium-coated microparticles (EP-A-486234) or microprojectile bombardment to induce wounding followed by co-cultivation with Agrobacterium (EP-A-486233).

[0047] Following transformation, a plant may be regenerated, e.g., from single cells, callus tissue or leaf discs, as is standard in the art. Almost any plant can be entirely regenerated from cells, tissues, and organs of the plant. Available techniques are reviewed in Vasil et al. (1984) in Cell Culture and Somatic Cell Genetics of Plants, Vols. I, II, and III, Laboratory Procedures and Their Applications (Academic press); and Weissbach et al. (1989) Methods for Plant Mol. Biol.

[0048] The transformed plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting hybrid having expression of the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that expression of the desired phenotypic characteristic is stably maintained and inherited, and then seeds harvested to ensure expression of the desired phenotypic characteristic has been achieved.

[0049] The particular choice of a transformation technology will be determined by its efficiency to transform certain plant species as well as the experience and preference of the person practicing the invention with a particular methodology of choice. It will be apparent to the skilled person that the particular choice of a transformation system to introduce nucleic acid into plant cells is not essential to or a limitation of the invention, nor is the choice of technique for plant regeneration.

[0050] Also according to the invention, there is provided a plant cell having the constructs of the invention. A further aspect of the present invention provides a method of making such a plant cell involving introduction of a vector including the construct into a plant cell. For integration of the construct into the plant genome, such introduction will be followed by recombination between the vector and the plant cell genome to introduce the sequence of

nucleotides into the genome. RNA encoded by the introduced nucleic acid construct may then be transcribed in the cell and descendants thereof, including cells in plants regenerated from transformed material. A gene stably incorporated into the genome of a plant is passed from generation to generation to descendants of the plant, so such descendants should show the desired phenotype.

[0051] The present invention also provides a plant comprising a plant cell as disclosed. Transformed seeds and plant parts are also encompassed. As used herein, the expressions "cell," "cell line," and "cell culture" are used interchangeably and all such designations include progeny. Thus, the words "transformants" and "transformed cells" include the primary subject cell and cultures derived therefrom without regard for the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to naturally occurring, deliberate, or inadvertent caused mutations. Mutant progeny that have the same function or biological activity as screened for in the originally transformed cell are included. Where distinct designations are intended, it will be clear from the context.

[0052] In addition to a plant, the present invention provides any clone of such a plant, seed, or hybrid descendants, and any part of any of these, such as cuttings or seed. The invention provides any plant propagule that is any part which may be used in reproduction or propagation, sexual or asexual, including cuttings, seed, and so on. Also encompassed by the invention is a plant which is a sexually or asexually propagated off-spring, clone, or descendant of such a plant; or any part or propagule of said plant, off-spring, clone, or descendant. Plant extracts and derivatives are also provided.

[0053] As is apparent to the routineer in this technology, the disclosed methods allow for the expression of a gene of interest in any plant. The invention thus relates generally to methods for the production of transgenic plants (both monocots and dicots). As used herein, the term "transgenic plants" refers to plants (algae, monocots, or dicots), comprising plant cells in which homologous or heterologous polynucleotides are expressed as the result of manipulation by the hand of man.

[0054] As is apparent to one of ordinary skill in the art, the peptides encoded by the disclosed herein may be encoded by multiple polynucleotide sequences because of the redundancy of the genetic code. It is well within the skill of a person trained in the art to create these alternative DNA sequences encoding the same, or essentially the same, amino acid sequences. These variant DNA sequences are within the scope of the subject invention.

[0055] The terms "purified" and "isolated", when referring to a polynucleotide, nucleotide, or nucleic acid, indicate a nucleic acid the structure of which is not identical to that of any naturally occurring nucleic acid or to that of any fragment of a naturally occurring genomic nucleic acid spanning more than three separate genes. The term therefore covers, for example, (a)

a DNA which has the sequence of part of a naturally occurring genomic DNA molecules but is not flanked by both of the coding or non-coding sequences that flank that part of the molecule in the genome of the organism in which it naturally occurs (e.g., DNA excised with a restriction enzyme); (b) a nucleic acid incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any naturally occurring vector or genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment; and (d) a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. Specifically excluded from this definition are nucleic acids present in mixtures of (i) DNA molecules, (ii) transfected cells, and (iii) cell clones, e.g., as these occur in a DNA library such as a cDNA or genomic DNA library.

[0056] The term "polynucleotide" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule and thus includes double- and single-stranded DNA and RNA. It also includes known types of modifications, for example, labels which are known in the art, methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications, such as those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as proteins (including for e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide.

[0057] "Control elements" include both "transcriptional control elements" and "translational control elements". "Transcriptional control elements" include "promoter", "enhancer", and "transcription termination" elements. Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription [Maniatis et al. [1987] Science 236:1237). Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in plants, yeast, insect and mammalian cells and viruses (analogous control elements, i.e., promoters, are also found in prokaryotes). The selection of a particular promoter and enhancer depends on what cell type is to be used to express the peptide of interest. Some eukaryotic promoters and enhancers have a broad host range while others are functional in a limited subset of cell types [for review see Voss et al. [1986] Trends Biochem. Sci. 11:287 and Maniatis et al. [1987] supra. Transcriptional control elements suitable for use in plants are well known in the art. "Translational control elements" include translational initiation regions and translational termination regions functional in plants.

[0058] A number of promoters can be used in the practice of the invention. The promoters can be selected based on the desired outcome. Strong promoters may be used to produce high levels of gene transcription. Alternatively, inducible promoters may be used to selectively active gene transcription when the appropriate signal is provided. Constitutive promoters may be utilized to continuously drive gene transcription. Tissue specific promoters may also be used in the practice of the invention in order to provide localized production of gene transcripts in a desired tissue. Developmental promoters may, likewise, be used to drive transcription of a gene during a particular developmental stage of the plant. Thus, a gene of interest can be combined with constitutive, tissue-specific, inducible, developmental, or other promoters for expression in plants depending upon the desired outcome.

[0059] Constitutive promoters include, for example, CaMV 35S promoter (Odell et al. (1985) Nature 313:810-812; rice actin (McElroy et al. (1990) Plant Cell 2:163-171; ubiquitin (Christensen et al. (1989) Plant Mol. Biol. 12:619-632 and Christensen et al. (1992) Plant Mol. Biol. 18:675-689); pEMU (Last et al. (1991) Theor. Appl. Genet. 81:581-588); MAS (Velten et al. (1984) EMBO J. 3:2723-2730); ALS promoter (U. S. Patent No. 5,659,026), and the like. Other constitutive promoters include those in U. S. Patent Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; and 5,608,142. Each of the aforementioned patents and references is hereby incorporated by reference in its entirety.

[0060] A number of inducible promoters are known in the art. For example, a pathogen-inducible promoter can be utilized. Such promoters include those from pathogenesis-related proteins (PR proteins), which are induced following infection by a pathogen; e.g., PR proteins, SAR proteins, beta-1,3-glucanase, chitinase, etc. See, for example, Redolfi et al. (1983) Neth. J. Plant Pathol. 89:245-254; Uknes et al. (1992) Plant Cell 4:645-656; and Van Loon (1985) Plant Mol. Virol. 4:111-116; Marineau et al. (1987) Plant Mol. Biol. 9:335-342; Matton et al. (1989) Molecular Plant-Microbe Interactions 2:325-331; Somsisch et al. (1986) Proc. Natl. Acad. Sci. USA 83:2427-2430; Somsisch et al. (1988) Mol. Gen. Genet. 2:93-98; and Yang (1996) Proc. Natl. Acad. Sci. USA 93:14972-14977. See also, Chen et al. (1996) Plant J. 10:955-966; Zhang et al. (1994) Proc. Natl. Acad. Sci. USA 91:2507-2511; Warner et al. (1993) Plant J. 3:191-201; Siebertz et al. (1989) Plant Cell 1:961-968; U. S. Patent No. 5,750,386; Cordero et al. (1992) Physiol. Mol. Plant Path. 41:189-200; each of which is incorporated by reference in its entirety.

[0061] Wound-inducible promoters may be used in the genetic constructs of the invention. Such wound-inducible promoters include potato proteinase inhibitor (pin II) gene (Ryan (1990) Ann. Rev. Phytopath. 28:425-449; Duan et al. (1996) Nature Biotechnology 14:494-498; wun1 and wun2, U. S. Patent No. 5,428,148; win1 and win2 (Stanford et al. (1989) Mol. Gen. Genet. 215:200-208); systemin (McGurl et al. (1992) Science 225:1570-1573); WIP1 (Rohmeier et al. (1993) Plant Mol. Biol. 22:783-792; Eckelkamp et al. (1993) FEBS Letters

323:73-76); MPI gene (Corderok et al. (1994) Plant J. 6(2):141-150; and the like. These references are also incorporated by reference in their entireties.

[0062] Tissue specific promoters can also be used in the practice of the subject invention. For example, leaf-specific promoters can similarly be used if desired, and are taught in references which include Yamamoto et al. (1997) Plant J. 12(2):255-265; Kawamata et al. (1997) Plant Cell Physiol. 38(7):792-803; Hansen et al. (1997) Mol. Gen. Genet. 254(3):337-343; Russel et al. (1997) Transgenic Res. 6(2):157-168; Rinehart et al. (1996) Plant Physiol. 112(3):1331-1341; Van Camp et al. (1996) Plant Physiol. 112(2):525-535; Canevascini et al. (1996) Plant Physiol. 112(2):513-524; Yamamoto et al. (1994) Plant Cell Physiol. 35(5)773-778; Lam (1994) Results Probl. Cell Differ. 20:181-196; Orozco et al. (1993) Plant Mol. Biol. 23(6):1129-1138; Matsuoka et al. (1993) Proc. Natl. Acad. Sci USA:90(20) 9586-9590; and Guevara-Garcia et al. (1993) Plant J. 4(3):495-505. Alternatively, root-specific promoters are known and can be selected from the many available from the literature. See, for example, Hire et al. (1992) Plant Mol. Biol. 20(2):207-218 (soybean root-specific glutamine synthetase gene); Keller and Baumgartner (1991) Plant Cell 3(10):1051-1061 (root-specific control element in the GRP 1.8 gene of French bean); Sanger et al. (1990) Plant Mol. Biol. 14(3):433-443 (root-specific promoter of the mannopine synthase (MAS) gene of Agrobacterium tumefaciens) Miao et al. (1991) Plant Cell 3(1):11-22 (full-length cDNA clone encoding cytosolic glutamine synthetase (GS), which is expressed in roots and root nodules of soybean). Bogusz et al. (1990) Plant Cell 2(7):633-641 (root specific promoters from hemoglobin genes from the nitrogen-fixing nonlegume Parasponia andersonii and the related non-nitrogen-fixing nonlegume Trema tomeniosa; Leach and Aoyagi (1991) Plant Science (Limerick) 79(1):69-76 (rolC and rolD root-including genes of Agrobacterium rhizogenes); Teeri et al. (1989) EMBO J.. 8(2):343-350 (octopine synthase and TR2' gene); (VfENOD-GRP3 gene promoter); Kuster et al. (1995) Plant Mol. Biol. 29(4):759-772 and Capana et al. (1994) Plant Mol. Biol. 25(4):681-691 (rolB promoter). See also U. S. Patent Nos. 5,837,876; 5,750,386; 5,633,363; 5,459;252; 5,401,836; 5,110,732; and 5,023,179.

[0063] Other tissue specific promoters can also be used in the practice of the subject invention (see, for example U.S. Patent No. 6,544,783). For example, xylem/vascular/tracheid-specific promoters, such as those disclosed in Milioni et al. (2002) Plant Cell, 14:2813-2824; Zhong et al. (1999) Plant Cell, 11:2139-2152; Ito et al. (2002) Plant Cell, 14:3201-3211; Parker et al. (2003) Development 130:2139-2148; Bourquin et al. (2002) Plant Cell 14:3073-3088 (each of which is hereby incorporated by reference in its entirety) can be used in the practice of the subject invention.

[0064] "Seed-preferred" promoters include both "seed-specific" promoters (those promoters active during seed development such as promoters of seed storage proteins) as well as "seed-germinating" promoters (those promoters active during seed germination). See Thompson

et al. (1989) Bioassays 10:108, herein incorporated by reference. Such seed-preferred promoters include, but are not limited to, Cim1 (cytokinin-induced message); cZ10B1 (Maize 19 kDa zein); celA (cellulose synthase); gama-zein; Glob-1; bean β-phaseolin; napin; β-conglycinin; soybean lectin; cruciferin; maize 15 kDa zein; 22 kDa zein; 27 kDa zein; g-zein; waxy; shrunken 1; shrunken 2; globulin 1; etc.

[0065] "Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

[0066] As used herein, the term "expression cassette" refers to a molecule comprising at least one coding sequence operably linked to a control sequence which includes all nucleotide sequences required for the transcription of cloned copies of the coding sequence and the translation of the mRNAs in an appropriate host cell. Such expression cassettes can be used to express eukaryotic genes in a variety of hosts such as bacteria, green algae, cyanobacteria, plant cells, fungal cells, yeast cells, insect cells and animal cells. Under the invention, expression cassettes can include, but are not limited to, cloning vectors, specifically designed plasmids, viruses or virus particles. The cassettes may further include an origin of replication for autonomous replication in host cells, selectable markers, various restriction sites, a potential for high copy number and strong promoters.

[0067] By "vector" is meant any genetic element, such as a plasmid, phage, transposon, cosmid, chromosome, virus etc., which is capable of replication when associated with the proper control elements and which can transfer gene sequences between cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

[0068] During the preparation of the constructs, the various fragments of DNA will often be cloned in an appropriate cloning vector, which allows for amplification of the DNA, modification of the DNA or manipulation of the DNA by joining or removing sequences, linkers, or the like. Preferably, the vectors will be capable of replication to at least a relatively high copy number in E. coli. A number of vectors are readily available for cloning, including such vectors as pBR322, vectors of the pUC series, the M13 series vectors, and pBluescript vectors (Stratagene; La Jolla, Calif.).

[0069] In order to provide a means of selecting transformed plants or plant cells, the vectors for transformation will typically contain a selectable marker gene. Marker genes are expressible DNA sequences which express a polypeptide which resists a natural inhibition by, attenuates, or inactivates a selective substance. Examples of such substances include antibiotics and, in the case of plant cells, herbicides. Selectable markers for use in animal, bacterial, plant, fungal, yeast, and insect cells are well known in the art. Exemplary selectable markers include

bacterial transposons Tn5 or Tn 601(903) conferring resistance to aminoglycosides (selection for Geneticin-resistance (G418R), mycophenolic acid resistance (MPAR) (utilizing E. coli guanosine phosphoribosyl transferase (gpt) encoding the enzyme XGPRT; selection is performed on medium containing MPA and xanthin), methotrexate resistance (MTXR), or cadmium-resistance which incorporates the mouse metallotheionein gene (as cDNA cassette) on the vector which detoxifies heavy metal ions by chelating them.

[0070] Alternatively, a marker gene may provide some visible indication of cell transformation. For example, it may cause a distinctive appearance or growth pattern relative to plants or plant cells not expressing the selectable marker gene in the presence of some substance, either as applied directly to the plant or plant cells or as present in the plant or plant cell growth media. The use of such a marker for identification of plant cells containing a plastid construct has been described (Svab et al. [1993] supra). Numerous additional promoter regions may also be used to drive expression of the selectable marker gene, including various plant promoters and bacterial promoters which have been shown to function in plants.

[0071] A number of other markers have been developed for use with plant cells, such as resistance to chloramphenicol, the aminoglycoside G418, hygromycin, or the like. Other genes which encode a product involved in chloroplast metabolism may also be used as selectable markers. For example, genes which provide resistance to plant herbicides such as glyphosate, bromoxynil or imidazolinone may find particular use. Such genes have been reported (Stalker et al. [1985] J. Biol. Chem. 260:4724-4728 (glyphosate resistant EPSP); Stalker et al. [1985] J. Biol. Chem. 263:6310-6314 (bromoxynil resistant nitrilase gene); and Sathasivan et al. [1990] Nucl. Acids Res. 18:2188 (AHAS imidazolinone resistance gene)).

[0072] Another aspect of the invention provides vectors for the cloning and/or the expression of a polynucleotide sequences taught herein in procaryotic or animal cells. The subject invention also provides for the expression of a polypeptide, peptide, derivative, or variant encoded by a polynucleotide sequence disclosed herein comprising the culture of a procaryotic or animal cell (a host cell) transformed with a polynucleotide of the subject invention under conditions that allow for the expression of a polypeptide, biologically active fragment, or multimeric construct encoded by said polynucleotide and, optionally, recovering the expressed polypeptide, peptide, derivative, or analog.

[0073] In this aspect of the invention, the polynucleotide sequences can be regulated by a second nucleic acid sequence so that the protein or peptide is expressed in a host cell transformed with the recombinant DNA molecule. For example, expression of a protein or peptide may be controlled by any promoter/enhancer element known in the art. Promoters which may be used to control expression include, but are not limited to, the CMV-IE promoter, the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3' long

terminal repeat of Rous sarcoma virus (Yamamoto, et al., 1980, Cell 22:787-797), the herpes simplex thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42); prokaryotic vectors containing promoters such as the β-lactamase promoter (Villa-Kamaroff, et al., 1978, Proc. Natl. Acad. Sci. U.S.A. 75:3727-3731), or the tac promoter (DeBoer, et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:21-25); see also "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242:74-94; plant expression vectors comprising the nopaline synthetase promoter region (Herrera-Estrella et al., 1983, Nature 303:209-213) or the cauliflower mosaic virus 35S RNA promoter (Gardner, et al., 1981, Nucl. Acids Res. 9:2871), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella et al., 1984, Nature 310:115-120); promoter elements from yeast or fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, and/or the alkaline phosphatase promoter.

[0074] The vectors according to the invention are, for example, vectors of plasmid or viral origin. In a specific embodiment, a vector is used that comprises a promoter operably linked to a nucleic acid sequence encoding a polypeptide as disclosed herein, one or more origins of replication, and, optionally, one or more selectable markers (e.g., an antibiotic resistance gene). Expression vectors comprise regulatory sequences that control gene expression, including gene expression in a desired host cell. Exemplary vectors for the expression of the polypeptides of the invention include the pET-type plasmid vectors (Promega) or pBAD plasmid vectors (Invitrogen) or those provided in the examples below. Furthermore, the vectors according to the invention are useful for transforming host cells so as to clone or express the polynucleotide sequences of the invention.

[0075] The invention also encompasses the host cells transformed by a vector according to the invention. These cells may be obtained by introducing into host cells a nucleotide sequence inserted into a vector as defined above, and then culturing the said cells under conditions allowing the replication and/or the expression of the polynucleotide sequences of the subject invention.

[0076] The host cell may be chosen from eukaryotic or prokaryotic systems, such as for example bacterial cells, (Gram negative or Gram positive), yeast cells (for example, Saccharomyces cereviseae or Pichia pastoris), animal cells (such as Chinese hamster ovary (CHO) cells), plant cells (e.g., algae), and/or insect cells using baculovirus vectors. In some embodiments, the host cells for expression of the polypeptides include, and are not limited to, those taught in U.S. Patent Nos. 6,319,691, 6,277,375, 5,643,570, or 5,565,335, each of which is incorporated by reference in its entirety, including all references cited within each respective patent.

[0077] Furthermore, a host cell strain may be chosen which modulates the expression of

the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain inducers; thus, expression of the genetically engineered polypeptide may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (e.g., glycosylation, phosphorylation) of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce an unglycosylated core protein product. Expression in yeast will produce a glycosylated product. Expression in mammalian cells can also to provide glycosylation of a protein.

[0078] The subject invention provides one or more isolated polypeptides comprising:

(a) SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47;

(b) a heterologous polypeptide sequence fused, in frame, to a polypeptide comprising SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47;

(c) a fragment of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47, wherein said fragment exhibits at least one biological function of the polypeptide of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47; or

(d) a variant having at least 70% homology to a polypeptide comprising SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47, wherein said variant exhibits at least one biological function of the polypeptide comprising SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47.

[0079] The term "peptide" may be used interchangeably with "oligopeptide" or "polypeptide" in the instant specification to designate a series of residues, typically L-amino acids, connected one to the other, typically by peptide bonds between the α-amino and carboxyl groups of adjacent amino acids. Linker elements can be joined to the polypeptides of the subject invention through peptide bonds or via chemical bonds (e.g., heterobifunctional chemical linker elements).

[0080] The subject invention encompasses polypeptide fragments of the full-length polypeptides disclosed herein. Polypeptide fragments, according to the subject invention, usually comprise a contiguous span of at least 5 consecutive (or contiguous) amino acids. The maximum length for a polypeptide fragment in the context of this invention is an integer that is one amino acid less than the full length of a particular SEQ ID NO: from which the fragment was derived. In certain preferred embodiments, fragments of the polypeptides of the subject invention retain at least one biological activity/function of the full-length polypeptide from which they are derived

(e.g., such similar or identical enzymatic activity or the ability to provide resistance to an antibiotic or transport an antibiotic out of a cell (see, for example, Table 4).

[0081] A "variant" polypeptide (or polypeptide variant) is to be understood to designate polypeptides exhibiting, in relation to the natural polypeptide, certain modifications. These modifications can include a deletion, addition, or substitution of at least one amino acid, a truncation, an extension, a chimeric fusion, a mutation, or polypeptides exhibiting post-translational modifications. Among the homologous polypeptides, those whose amino acid sequences exhibit between at least (or at least about) 70.00% to 99.99% (inclusive) identity to the full length, native, or naturally occurring polypeptide are another aspect of the invention. The aforementioned range of percent identity is to be taken as including, and providing written description and support for, any fractional percentage, in intervals of 0.01%, between 70.00% and, up to, including 99.99%. These percentages are purely statistical and differences between two polypeptide sequences can be distributed randomly and over the entire sequence length. Thus, variant polypeptides can have 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent identity with the polypeptide sequences of the instant invention. In certain preferred embodiments, variants of the polypeptides of the subject invention retain at least one biological activity/function of the full-length polypeptide from which they are derived (e.g., such as similar or identical enzymatic activity or the ability to provide resistance to an antibiotic or transport an antibiotic out of a cell (see, for example, Table 4).

[0082] Variant polypeptides can also comprise one or more heterologous polypeptide sequences (e.g., tags that facilitate purification of the polypeptides of the invention (see, for example, U.S. Patent No. 6,342,362, hereby incorporated by reference in its entirety; Altendorf et al. [1999-WWW, 2000] "Structure and Function of the Fo Complex of the ATP Synthase from Escherichia Coli," J. of Experimental Biology 203:19-28, The Co. of Biologists, Ltd., G.B.; Baneyx [1999] "Recombinant Protein Expression in Escherichia coli," Biotechnology 10:411-21, Elsevier Science Ltd.; Eihauer et al. [2001] "The FLAG™ Peptide, a Versatile Fusion Tag for the Purification of Recombinant Proteins," J. Biochem Biophys Methods 49:455-65; Jones et al. [1995] J. Chromatography 707:3-22; Jones et al. [1995] "Current Trends in Molecular Recognition and Bioseparation," J. of Chromatography A. 707:3-22, Elsevier Science B.V.; Margolin [2000] "Green Fluorescent Protein as a Reporter for Macromolecular Localization in Bacterial Cells," Methods 20:62-72, Academic Press; Puig et al. [2001] "The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification," Methods 24:218-29, Academic Press; Sassenfeld [1990] "Engineering Proteins for Purification," TibTech 8:88-93; Sheibani [1999] "Prokaryotic Gene Fusion Expression Systems and Their Use in Structural and Functional Studies of Proteins," Prep. Biochem. & Biotechnol. 29(1):77-90, Marcel Dekker, Inc.; Skerra et al. [1999] "Applications of a Peptide Ligand for Streptavidin: the Strep-

tag", Biomolecular Engineering 16:79-86, Elsevier Science, B.V.; Smith [1998] "Cookbook for Eukaryotic Protein Expression: Yeast, Insect, and Plant Expression Systems," The Scientist 12(22):20; Smyth et al. [2000] "Eukaryotic Expression and Purification of Recombinant Extracellular Matrix Proteins Carrying the Strep II Tag", Methods in Molecular Biology, 139:49-57; Unger [1997] "Show Me the Money: Prokaryotic Expression Vectors and Purification Systems," The Scientist 11(17):20, each of which is hereby incorporated by reference in their entireties), or commercially available tags from vendors such as such as STRATAGENE (La Jolla, CA), NOVAGEN (Madison, WI), QIAGEN, Inc., (Valencia, CA), or InVitrogen (San Diego, CA). Alternatively, the heterologous sequences may provide for the multimerization of the polypeptides of the subject invention (see, e.g., US Patent Number 5,478,925, WO 98/49305, or U.S. Pat. No. 5,073,627, Landschulz *et al.*, (1988), Science. 240:1759, WO 94/10308, Hoppe *et al.*, (1994), FEBS Letters. 344:191). Other methods of making multimers include the addition of cysteine or biotin to the C-terminus or N-terminus of the polypeptide using techniques known in the art. Where biotin is attached to a polypeptide, avidin can be utilized to create multimers of the polypeptides to which the biotin element is attached (see, e.g., US Patent Number 5,478,925 for numerous methods of multimerization). Multimers of the invention may also be generated using chemical or genetic engineering techniques known in the art.

[0083] The invention, thus, provides a novel antibiotic family, Albicidins, produced by three novel biosynthetic gene clusters (XALB1, XALB2, and XALB3) contained within a host cell DNA in which one strand comprises non-contiguously SEQ. ID No. 1, SEQ. ID No. 2 and SEQ ID No. 3, and the cell expresses the DNA to provide peptides including those named AlbI (SEQ ID No. 26) (encoded by SEQ ID No. 20), AlbII (SEQ ID No. 27) (encoded by SEQ ID No. 21), AlbIII (SEQ ID No. 28) (encoded by SEQ ID No. 22), AlbIV (SEQ ID No. 29) (encoded by SEQ ID No. 23), AlbVI (SEQ ID No. 31) (encoded by SEQ ID No. 18), AlbVII (SEQ ID No. 32) (encoded by SEQ ID No. 17), AlbVIII (SEQ ID No. 33) (encoded by SEQ ID No. 16), AlbIX (SEQ ID No. 34) (encoded by SEQ ID No. 15), AlbX (SEQ ID No. 35) (encoded by SEQ ID No. 10), AlbXI (SEQ ID No. 36) (encoded by SEQ ID No. 9), AlbXII (SEQ ID No. 37) (encoded by SEQ ID No. 8), AlbXIII (SEQ ID No. 38) (encoded by SEQ ID No. 7), AlbXIV (SEQ ID No. 39) (encoded by SEQ ID No. 6), AlbXV (SEQ ID No. 40) (encoded by SEQ ID No. 5), AlbXVII (SEQ ID No. 42) (encoded by SEQ ID No. 11), AlbXVIII (SEQ ID No. 43) (encoded by SEQ ID No. 12), AlbXIX (SEQ ID No. 44) (encoded by SEQ ID No. 13), AlbXX (SEQ ID No. 45) (encoded by SEQ ID No. 14), AlbXXI (SEQ ID No. 46) (encoded by SEQ ID No. 24), and AlbXXII (SEQ ID No. 47) (encoded by SEQ ID No. 25), that in turn interact within the host cell to produce one or more antibiotics as more fully illustrated in Figure 11.

[0084] In one embodiment, the invention comprises a plurality of isolated and purified DNA strands which comprise nucleotide sequences selected from the group consisting of SEQ ID

No: 1 to SEQ. ID No. 25, each individual sequence, except the transposases AlbV (SEQ ID No. 30) (encoded by SEQ ID No. 19) and AlbXVI (SEQ ID No. 41) (encoded by SEQ ID No. 4) found in the XALB1 cluster, being necessary to the biosynthesis of the novel family of antibiotics, Albicidins.

[0085] The invention also includes the peptides or proteins encoded by the genes of the biosynthetic complex expressed by the combination of DNA with a strand having sequences SEQ ID Nos. 1 to 3. Proteins are named with roman numerals and the prefix Alb from AlbI to Alb XXII have the amino acid sequences of SEQ ID Nos. 26 to 47 (not in Roman numeral order but in the order of placement of the genes within sequences SEQ ID Nos. 1 to 3 that express each protein). Expression of the peptides having the amino acid sequences of SEQ ID Nos. 26 to 29, 31 to 40 and 42 to 47, have been found to be all required for the successful biosynthesis of Albicidins.

[0086] The invention further provides a method for producing Albicidins comprising providing a modified host cell with a heterologous DNA Albicidin Biosynthetic Gene Cluster or set of genes defined as DNA operably comprising DNA sequences substantially similar to SEQ ID Nos. 1 to 3. Substantially the same means DNA having sufficient homology to provide expressed proteins that function to provide an antibiotic material having the structural components identified herein. Preferably a given sequence will have at least 70 percent homology to one of SEQ ID Nos. 1 to 3, preferably 85% homology and most preferably at least 95% homology. The method includes the steps of modifying the DNA of the host cell to comprise an operable expression system for maintaining the modified host cell under conditions supporting biosynthesis of Albicidins and isolation of Albicidins from the host cell or its environment. The invention further provides a method of production of a group of novel antibiotic materials utilizing at least three of the Sequences selected from the group consisting of DNA SEQ ID No. 1 to SEQ ID No. 25 (excluding transposases encoded by SEQ ID Nos. 4 and 19) inclusive in combination with additional sequences to produce a modified Albicidin-like material.

[0087] More specifically, the invention provides DNA Sequences comprising at least about 68,498 base pairs and including an about 55,839 bp region from the genome of *X. albilineans* designated as XALB1 (Albicidin Biosynthetic Gene Cluster 1; SEQ ID. No. 1) an additional non-contiguous region of about 2,986 bp, XALB2 (Albicidin Biosynthetic Gene Cluster 2; SEQ ID. No. 2), and a third region of about 9,673 bp, XALB3 (Albicidin Biosynthetic Gene Cluster 3; SEQ ID. No. 3). Albicidin Biosynthetic Gene Clusters 1-3 may be referred to, collectively, as the Albicidin Biosynthetic Gene Clusters and these sequences were found to be required for biosynthesis of Albicidins. Homology analysis revealed the presence of (i) four large genes with a modular architecture characteristic of polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS) potentially involved in albicidin precursor biosynthesis; (ii) four

smaller genes potentially involved in albicidin substrate biosynthesis (iii) four modifying genes; (iv) one enzyme activating gene, (v) two regulatory genes, (vi) one chaperone gene, (vii) two genes of unknown function; and (viii) two resistance genes. These are named and discussed more fully below. Together these genes allow the successful operation of the biosynthetic pathway when cloned into suitable host cells.

[0088] Alignment of individual NRPS and PKS domains revealed an extraordinary biosynthetic apparatus believed to involve a *trans*-action of separate PKS and NRPS domains which could contribute to the production of multiple, structurally related albicidins by the same gene cluster. Furthermore, analysis of selectivity-conferring residues indicated that four NRPS modules of XALB1 specify an unusual substrate.

[0089] In an alternate embodiment the invention provides a method of producing a polyketide carrying para-aminobenzoic acid and/or carbamoyl benzoic acid by inserting at least one DNA fragment that encodes a PKS protein into a cell and causing the cell to express the encoded PKS protein under conditions such that the PKS protein functions to produce a polyketide carrying either a para-aminobenzoic acid or a carbamoyl benzoic acid or both. Another embodiment provides a method of producing polyketide/peptides carrying para-aminobenzoic acid and/or carbamoyl benzoic acid by inserting at least one DNA fragment that encodes a PKS protein into a cell and causing the cell to express the encoded PKS protein under conditions such that the PKS protein functions to produce a polyketide carrying either a para-aminobenzoic acid or a carbamoyl benzoic acid or both. In yet another embodiment, the invention provides a method of activating nonproteinogenic amino acids like paraminobenzoic acid and/or carbamoyl benzoic acid for incorporation into peptides or polyketides by inserting at least one DNA fragment that encodes a PKS protein into a cell and causing the cell to express the encoded PKS protein under conditions such that the PKS protein functions to produce a polyketide carrying either a para-aminobenzoic acid or a carbamoyl benzoic acid or both.

[0090] There are three regions of the *X. albilineans* genome specifying albicidin production. XALB2 and XALB3 regions each contain only one gene, both of which are required for post-translational activation and folding of albicidin PKS and NRPS enzymes. The XALB1, XALB2 and XALB3 gene clusters are characterized by an unusual hybrid NRPS-PKS system, indicating that albicidin biosynthesis may provide an excellent model for investigating the biosynthesis of hybrid polyketide-polypeptide metabolites in bacteria. The availability of three genomic regions involved in albicidin production, XALB1 and XALB2 and XALB3, also offers the ability to express individually the enzymes of the albicidin family biosynthetic pathway including structural, resistance, secretory and regulatory elements, and to engineer overproduction of albicidin in mutated or modified host cells of the invention. The invention overcomes prior art limitations in albicidin production due to low yields of toxin production in *X. albilineans* and may

also allow characterization of the chemical structure of albicidin as well as application of this potent inhibitor of prokaryote DNA replication.

[0091] The invention results from a number of unpredictable results namely the number and complexity of the enzymes involved in biosynthesis. The discovery of the complete sequence required for biosynthesis of Albicidins is previously unreported. The invention provides for a novel process for production of molecule having a polyketide-polypeptide backbone and the formula $C_{40}H_{35}O_{15}N_6$, a molecular weight of 839, and the structural elements shown in Figure 11.

[0092] The invention further includes (a) the Albicidin Family Biosynthetic Gene Cluster including (b) the structural and regulatory elements of the operons that encode c) the enzymes PKS-1, PKS-2, PKS-3, PKS-4, NRPS-1, NRPS-2, NRPS-3, NRPS-4, NRPS-5, NRPS-6 and NRPS-7 as well as (d) the proteins AlbI to AlbXXII, (e) the isolated enzymes, proteins, and active forms thereof, as well as mutants, fragments, and fusion proteins comprising any of the forgoing; (f) the uses of the enzymes or proteins encoded by the Albicidins Biosynthesis Gene Cluster or any one of its operons, (g) a host cell expressing one or more enzymes or proteins encoded by the Albicidin Family Biosynthetic Gene Cluster; (h) use of host cells having the Albicidins Biosynthesis Gene Cluster to produce an antibiotic; (i) methods of modifying the DNA sequences to produce members of a series of antibiotic compounds having structures related to Albicidins; (j) DNA sequences that encode the same proteins as any of SEQ. ID. Nos. 1 to 25 but differ in specific codons due to the multiplicity of codons that lead to expression of the same amino acid; (k) antibiotics produced by the process of expression of the Albicidin Family Biosynthetic Genes in a genetically modified host cell sustained in a culture medium and thereafter separation of the antibiotic from the host cell and culture medium; (l) an isolated and purified antibiotic produced by a process that includes at least three proteins coded by DNA sequences selected for the group consisting of SEQ. ID Nos. 1 to 25 in combination with additional enzymes that modify the product to provide a non-naturally occurring Albicidins like product having at least one of the useful properties reported for albicidin; and (m) a process for producing an antibiotic that comprises modifying a host cell to enhance expression of the DNA of the Albicidin Family Biosynthetic Gene Cluster by insertion of expression enhancing DNA into the genome of a *Xanthomonas albilineans* strain in a position operative to enhance expression of the enzymes of the Albicidin Family Biosynthetic Gene Cluster, culturing the modified host cell to produce an antibiotic and isolating the antibiotic. The products and methods described above have utility as proteins or as nucleic acids as the case may be, including such uses sources of pyrimidine or purine bases or amino acids, or as animal food supplements and the like, as well as the more important uses to provide antibiotics, plant disease treatment methods, genetically modified disease resistant plants, phytotoxins and the like.

[0093] The subject invention also provides an isolated and purified antibiotic produced

by a process that includes at least three proteins coded by the nucleic acids of the subject invention in combination with additional enzymes that modify the product to provide a non-naturally occurring Albicidin-like product having at least one of the useful properties reported for albicidin. In certain embodiments, the antibiotic or antibiotics have at least one of the general structures illustrated in Figure 11. In other embodiments, antibiotics of the subject invention have at least 4 of the structural elements illustrated in Figure 11, and an elemental composition of $C_{40}H_{35}N_6O_{15}$.

[0094] The invention further provides a method of protecting a plant against damage from albicidin that comprises applying an agent that blocks expression at least one gene in the Albicidin Biosynthetic Gene Clusters to the plant to be protected. Additional inventions include a method of obtaining agents useful in blocking expression of albicidin by screening materials against a modified host cell line that expresses the Albicidin Biosynthesis Gene Clusters and selecting for materials that stop or decrease albicidin production and a method of protecting a plant against phytotoxic damage from an antibiotic that comprises inserting into the plant and operably expressing at least one resistance gene from the Albicidin Biosynthesis Gene Clusters into the plant to be protected.

EXAMPLE 1 – Materials and Methods

[0095] Bacterial strains and plasmids. The source of bacterial strains and their relevant characteristics are described in Table 1.

[0096] Media, antibiotics, and culture conditions. *X. albilineans* strains were routinely cultured on modified Wilbrink=s (MW) medium at 30□C without benomyl (Rott *et al.*, 1994). For long-term storage, highly turbid distilled water suspensions of *X. albilineans* were supplemented with glycerol to 15% (vol/vol) and frozen at B80□C. For *X. albilineans*, MW medium was supplemented with the following antibiotics as required at the concentrations indicated: kanamycin, 10 or 25 μg/ml; and rifampicin, 50 μg/ml. *E. coli* strains were grown on Luria-Bertani (LB) agar or in LB broth at 37□C and were maintained and stored according to standard protocols (Sambrook *et al.*, 1989). For *E. coli*, LB medium was supplemented with the following antibiotics as required at the concentrations indicated: kanamycin, 50 μg/ml; ampicillin, 50 μg/ml.

[0097] Bacterial conjugation. DNA transfer between *E. coli* donor (DH5_MCR/pAlb389 or pAC389.1, Table 1) and rifampicin-resistant *X. albilineans* recipients (*X.* strains AM10, AM12, AM13, AM36 and AM37, Table 1) was accomplished by triparental conjugation with plasmid pRK2073 as the helper as described previously (Rott *et al.*, 1996).

[0098] Assay of albicidin production. Albicidin production was tested by a microbiological assay as described previously (Rott *et al.*, 1996). Rifampicin and kanamycin exconjugants were spotted with sterile toothpicks (2-mm-diameter spots) onto plates of SPA

medium (2% sucrose, 0.5% peptone, 1.5% agar) and incubated at 28° C for 2-5 days. The plates were then overlaid with a mixture of *E. coli* DH5α (10^7 cells in 2 ml of distilled water) plus 2 ml of molten 1.5% (wt/vol) Noble agar (Difco) at ca. 65° C and examined for inhibition zones after 24 h at 37C.

[0099] Nucleic acid manipulations. Standard molecular techniques were used to manipulate DNA (Sambrook *et al.*, 1989) except for total genomic DNA preparation. Total genomic DNA for southern blot hybridization was prepared as described by Gabriel and De Feyter (1992).

[00100] PCR Conditions. PCR amplifications were performed in an automated thermal cycler PTC-100™ (MJ Research, Inc). The 25 µl PCR reaction mix consisted of 100 ng of genomic DNA or 1 ng of plasmid DNA, 2.5 µl of 10X PCR buffer without MgCl2 (Eurobio), 80 µM dNTP mix, 2.5 units of EUROBIOTAQII®(Eurobio), 25 pmoles of each primer, 2.0 mM $MgCl_2$ (Eurobio) and sterilized distilled water to final volume. The PCR program was 95°C for 2 min, 25 cycles at 94°C for 1 min, Tm for 1 min and 72°C for 1 min, with a final 72°C extension for 5 min. Tm temperature was determined for each couple of primers and varied between 55°C and 60°C. A 5µl aliquot of each amplified product was analyzed by electrophoresis through a 1% agarose gel. For sequencing, PCR products were cloned with the pGEM®-T Easy Vector System (Promega).

[00100] Oligonucleotide synthesis. Oligonucleotides were purchased from Genome Express (Grenoble or Montreuil, France).

[00101] DNA sequencing. Automated DNA sequencing was carried out on double-stranded DNA by the dideoxynucleotide chain termination (Sanger *et al.*, 1977) using a Dye Terminator Cycle Sequencing kit and an ABI Perkin-Elmer sequencer according to the manufacturer's procedure. Both DNA strands were sequenced with universal primers or with internal primers (20mers). This service was provided by Genome Express (Grenoble, France). Computer-aided sequence analyses were carried out using Sequence Navigator™ (Applied Biosystems, Inc) and SeqMan (DNASTAR Inc.) programs.

[00102] Sequence analysis. Nucleotide sequences were translated in all six reading frames using EditSeq (DNASTAR Inc.). Potential products of ORFs longer than 100 b were compared to protein databases by the PSI-BLAST program (Swiss-Prot and Genbank) on the NCBI with site (ncbi.nlm.nih.gov/) using Altschul program (Altschul *et al.*, 1997). The TERMINATOR program of the Genetics Computer Group was used to identify putative Rho-independent transcription terminators.

[00103] **Procedures**

EXAMPLE 2 – Sequencing of the Double Strand Region of 55,839 Bp from *X. albilineans* Containing XALB1 SEQ ID NO. 1

[00104] In Figure 1 is presented a physical map and genetic organization of XALB1. In the figure, E and K are restriction endonuclease sites for *Eco*RI and *Kpn*I, respectively. Rectangular boxes represent DNA fragments labeled A through N. The numbers below each rectangular box are the number of Tn*5*-*gus* insertion sites previously located in each DNA fragment (Rott *et al.*, 1996). The DNA inserts carried by plasmids pALB571 and pALB540 are represented by bold bars above the physical map. The location and direction of putative orfs identified in the XALB1 gene cluster are shown by arrows. Precise positions and proposed functions for individual orfs are summarized in Tables 2 and 3, respectively. Position of insertional sites of eight albicidin-defective mutants determined by sequencing are indicated by vertical arrows. The location and direction of putative ORFS identified in the XALB1 gene cluster are shown by arrow shapes. These twenty putative ORFs are potentially organized in four or five operons, as indicated at the bottom of the figure. Patterns indicate NRPS and PKS genes (diagonal crosshatch), methyl transferase and esterase genes (hollow rectangles), carbamoyl transferase gene (fine crosshatch), benzoate-derived products biosynthesis genes (white), regulatory genes (vertical lined), resistance genes (diagonal lines) and other genes with function of unknown significance to albicidin production (black), and three insertional sites of eight albicidin-defective mutants determinated by sequencing are indicated by vertical arrows. Dotted regions in the physical map and in ORFs represent the two internal duplicated DNA regions of XALB1.

[00105] The sequence illustrated in Figure 1 was generated as follows. The sources of DNA are set out in Table 1. DNA fragments F, E, B, C, I, and G, generated by the digestion of cosmid pALB571 (Rott *et al.*, 1996) with *Eco*RI and/or *Kpn*I, were subcloned into pBCKS (+) and were sequenced from the resulting subclones, pBC/F, pBC/E, pBC/B, pBC/C, pBC/I and pBC/G. DNA fragment D' which corresponds to the part of fragment D present in cosmid pALB571 was sequenced from plasmid pUFR043/D' obtained following self ligation of the complete *Eco*RI digested cosmid pALB571. DNA fragment H was sequenced from pAM45.1 (Rott *et al.*, 1996), obtained following cloning into vector pBR325 of the 12kb *Eco*RI fragment carrying Tn*5* and flanking sequences from mutant strain XaAM45. DNA fragment A' contains the part of fragment A present in cosmid pALB571 and was subcloned into vector pBCKS (+) and the resulting plasmid pBC/A' was used for sequencing. The presence of a large internal duplication made alignment of sequence data obtained from pBC/A' difficult. This difficulty was

resolved using sequence data obtained from an additional plasmid, pAM4, obtained following cloning into vector pBluescript II KS (+) of the 12kb *Eco*RI fragment carrying Tn*5* and flanking sequences from mutant strain XaAM4, which contains only one copy of the large internal duplication. Sequence data from pBC/A' were used to determine the first 1542 bp of fragment A' between nucleotides C-19001 and G-20543. Sequence data from pAM4 and pBC/A' were used to determine the last 4823bp of fragment A' between nucleotides G-21653 and G-26477. The overlapping region between nucleotides G-20469 and C-22159 was amplified by PCR from cosmid pALB571 using primers contig13-1160 (5'gcgtaccgttgtccagtagg3') SEQ ID NO. 48 and pAM4-14 (5'gctggaaaccgagaatctga3') SEQ ID NO. 49, and was sequenced. Resulting sequence data were used to complete sequencing of DNA fragment A'. The junctions A/F, F/H, H/E, E/B, B/C, C/I, I/G, G/D between corresponding DNA fragments were sequenced directly from cosmid pALB571. EcoRI DNA fragment containing fragments A and F was subcloned from pALB540 into pBCKS (+), and the resulting plasmid pBC/AF was used to determine the part of DNA fragment A which was not present in cosmid pALB571 between nucleotides G-13682 and G-19001. EcoRI DNA fragments J, K, L, N were subcloned from pALB540 into pBCKS (+) and were sequenced from resulting plasmid pBC/J, pBC/K, pBC/L, and pBC/N. The junctions L/K, K/J and J/A between corresponding DNA fragments were sequenced directly from cosmid pALB540. DNA region between nucleotides G-7517 and T-8721 was amplified by PCR from cosmid pALB540 using primers E114 (5'gacacgatcagccgctagga3') SEQ ID NO. 50 and EI4-380 (5'accagcagttgggccagcct3') SEQ ID NO. 51 and was sequenced. Resulting sequence data were used to determine the sequence of fragment M and of junctions N/M and M/L. The nucleotide sequence of 55,839 bp containing the entire major gene cluster involved in Albicidin production was sequenced on both strands.

<u>EXAMPLE 3 – Analysis of the Large Internal Duplications in the DNA Sequence of XALB1</u>

[00106] The sequence of the 55,839 bp genomic region (SEQ ID NO. 1) contains two large internal duplications as shown by the dotted regions in the physical map of Figure 1. A direct duplication of 1736 bp was located in DNA fragment A between nucleotides G-19904 and G-21639 and between nucleotides G-23057 and G-24792. Another direct duplication of a 2727 bp was found in DNA fragments B and C between nucleotides C-40410 and G-43136 and between nucleotides C-46644 and G-49370. Comparison of the two copies of each duplication revealed that the two copies of the 1736 bp duplication are identical except for one nucleotide at position 21058, and that the two copies of the 2727 bp duplication are 98.8% identical and differ by 30 nucleotides.

EXAMPLE 4 – Comparison of XALB1 with the *xabB EcoRI* Fragment

[00107] Comparison of the DNA sequence of the 55,839 bp genomic region described in this study with the partial DNA sequence of 16,511 bp of the same region in Huang et al., 2001 (described by Huang et al. as an *Eco*RI fragment including full length xabB from *X. albilineans* strain Xa13 [GenBank accession No. AF239749]), revealed that the DNA sequence from strain Xa13 over 16,511 bp is identical to the sequence from strain Xa23R1, described herein, with the following exceptions: 1) five nucleotides are different at positions 42963, 42972, 42980, 43014 and 43071 of the XALB1 sequence, and 2) nucleotides from positions 43137 to 49370 are missing (internal to *albI*; refer Fig. 1). Analysis of genomic DNA of seven strains isolated from different countries (Australia, Reunion Island, Kenya, Zimbabwe and USA), digested by *Kpn*I and hybridized with the pBC/C plasmid (Table 1) labeled with ^{32}P, revealed that two DNA fragments corresponding to the XALB1 fragments B and C were present in all strains (data not shown). This result indicated that all studied strains contain *albI* and not *xabB* because in *albI* the pBC/C plasmid probe hybridizes with the large internal duplication present in both DNA fragments B and C (Figure 1). Based on this observation we postulated that the DNA sequence of XabB reported as full length by Birch in PCT WO 02/24736 A1 (Their seq. ID#1) appears to be incomplete and missing 6,234 bp of DNA sequence encoding 2,078 amino acids.

EXAMPLE 5 – Reading Frame Analysis in XALB1

[00108] Analysis of the 55,839 bp double strand region for coding sequences revealed the presence of 20 open reading frames (ORFs) designated *albI* to *albXX* (Table 2 below) which are distributed in four groups of genes according to their position and their orientation in the XALB1 cluster (Figure 1). Genes of each group may form part of the same operon as judged by their overlapping stop and start codons, or by the relatively short intergenic region which varies from 5 to 274 nucleotides. The 20 ORFs appear to be organized in four operons: operon 1 formed by *albI* - *albIV*; operon 2 by *albV* - *albIX*; operon 3 by *albX* - *albXVI*; operon 4 by *albXVII* - *albXX*. The majority of *alb* ORFs are initiated with an ATG codon, except *albI* and *albXVII* which are initiated with a TTG codon, and *albIV* and *albVI* which are initiated with a GTG start codon. In seven ORFs of XALB1, start codons are preceded by the consensus sequence GAGG which may correspond to the ribosome binding site. Other ORFs are preceded by a less conserved sequence which contain at least three nucleotides A or G and which may serve as a weak ribosome binding site.

EXAMPLE 6 – Sequencing of the *Tn5* insertional site of eight tox⁻ mutants previously located in XALB1

[00109] Eight of the 45 *X. albilineans* Tox⁻ mutants complemented by cosmid pALB540 and/or cosmid pALB571 and previously described (Rott *et al.*, 1996) were further analyzed. All eight mutants contain a single Tn*5* insertion and correspond to the following *X. albilineans* strains: XaAM7, XaAM15, XaAM45, and XaAM52 which are complemented by pALB571 but not by pALB540; XaAM4, XaAM29 and XaAM40 which are complemented by both cosmids; and XaAM1 which is complemented by pALB540 but not by pALB571. The Tn*5* insertional site of each Tox⁻ mutant was sequenced from plasmids obtained following cloning in pBR325 or in pBluescript II KS (+) of the EcoRI fragments carrying Tn*5* and flanking sequence using the sequencing primer GUSN (5'tgcccacaggccgtcgagt3') SEQ ID No. 52 that annealed 135 bp downstream from the insertional sequence IS50L of Tn*5*-*gus*A. The sequence of the Tn*5* insertional site was compared with the 55,839 bp sequence containing XALB1 in order to determine the *alb* gene disrupted in each Tox⁻ mutant. *albI* is disrupted by the Tn*5* insertion in XaAM15 and XaAM45 at position 33443 and 34229, respectively (Figure 1). *albIV* is disrupted by the Tn*5* insertion in XaAM7 and XaAM52 at position 53704 and 53915, respectively. *albIX* is disrupted by the Tn*5* insertion in XaAM4, XaAM29 and XaAM40 at position 21653, 23444 and 24376, respectively. *alb XI* is disrupted by the Tn*5* insertion in XaAM1 at position 13301. These results are in accordance with the previous characterization of Tox⁻ mutants using Southern blot hybridization (Rott *et al.*, 1996), except for XaAM1. The Tn*5*-*gusA* insertion site of XaAM1 was previously located in DNA fragment A (Rott *et al.*, 1996) but results of this study showed that this site is located in DNA fragment J (Figure 1).

EXAMPLE 7 – Homology analysis of proteins potentially encoded by XALB1

[00110] Preliminary functional assignments of individual ORFs were made by comparison of the deduced gene products with proteins of known functions in the Genbank database. The results are set out in Table 3 below. Among the ORFs identified from the sequenced XALB1 gene cluster, we found (i) four genes, *albI* SEQ ID No. 20, *albIV* SEQ ID No. 23, *albVII* SEQ ID No. 17 and *albIX* SEQ ID No. 15, encoding PKS and/or NRPS modules; (ii) one carbamoyl transferase gene, *albXV* SEQ ID No. 5; (iii) two esterase genes, *albXI* SEQ ID No. 9 and *albXIII* SEQ ID No. 7; (iv) two methyltransferase genes, *albII* SEQ ID No. 21 and *albVI* SEQ ID No. 18; (v) two benzoate-derived products biosynthesis genes, *albXVII* SEQ ID No. 11 and *albXX* SEQ ID No.14; (vi) two putative albicidin biosynthesis regulatory genes, *albIII* SEQ ID No. 22 and *albVIII* SEQ ID No. 16; (vii) two putative albicidin resistance genes, *albXIV* SEQ ID No. 6 and *albXIX* SEQ ID No. 13; and (viii) two additional ORFs encoding proteins similar to transposition proteins, *albV* SEQ ID No. 19 and *albXVI* SEQ ID No. 4. No known

function was found in the database for *albX* SEQ ID No. 10 and *albXII* SEQ ID No. 8. The potential product of *albXVIII* SEQ ID No. 12 appeared to be a truncation of an enzyme with strong similarity to 4-amino-4-deoxychorismate lyase and branched-chain amino acid aminotransferase. Since the gene encoding the predicted product is roughly half the length of other such lyase or aminotransferase genes, *albXVIII* may be the result of a recombination event and may be non functional.

EXAMPLE 8 – The *alb* PKS and/or NRPS genes

[00111] The potential product of *albI*, designated AlbI SEQ ID No. 20, is a protein of 6879 aa with a predicted size of 755.9 kDa. This protein is very similar to the potential product of the *xabB* gene from *X. albilineans* strain Xa13 from Australia (Huang *et al.*, 2001), but it differs in length and size (See Table 4 below). XabB is a protein of 4801 amino acids with a predicted size of 525.7 kDa. Comparison of AlbI with XabB revealed that the N-terminal regions from Met-1 to Ile-4325 of both proteins are identical except for five amino-acids which are Tyr-3941, Pro-3952, Ala-4054, Ala-4271 and Gln-4284 in AlbI and His-3941, Ala-3952, Val-4054, Val-4271 and Glu-4284 in XabB. The same comparison revealed that the AlbI C-terminal region from Arg-6404 to the stop codon is 100% identical to the XabB C-terminal region from Arg-4326 to the stop codon.

[00112] The N-terminal region (from Met-1 to Asp-3235) of AlbI is 100% identical to the corresponding region in XabB which was previously described as similar to many microbial modular PKS (Huang *et al.*, 2001). This PKS region may be divided into three modules (Figure 2). Abbreviations used in the Figure are: A, adenylation; ACP, acyl carrier protein; AL, acyl-CoA ligase; C, condensation; KR, ß-ketoacyl reductase; KS, ß-ketoacyl synthase; NRPS, nonribosomal peptide synthase; PCP, peptidyl carrier protein; PKS, polyketide synthase; TE, thioesterase; HBCL, 4-hydroxybenzoate-CoA ligase. The question mark in the NRPS-2 domain indicates that this A domain is incomplete. The first module designated PKS-1 contains acyl-CoA ligase (AL) and acyl carrier protein (ACP1) domains. The second module designated PKS-2 contains ß-ketoacyl synthase (KS1) and ß-ketoacyl reductase (KR) domains followed by two consecutive ACP domains (ACP2 and ACP3). The third module designated PKS-3 contains a KS domain (KS2) followed by a PCP domain (PCP1). Apart from their very high similarity with XabB, these three PKS modules exhibited the highest degree of overall similarity with polyketide synthases SafB and PksM from *Myxococcus xanthus* and *Bacillus subtilis*, respectively (Table 4). The motifs characteristic of these domains are 100% identical to those of XabB which were previously aligned with those from other organisms (Huang *et al.*, 2001). The AL domain contains the conserved adenylation core sequence (SGSSG) and the ATPase motif (TGD). The three ACP domains contain a 4'-phosphopantetheinyl-binding cofactor box GxDS(IL), except that

A replaced G in ACP1. Both KS domains contain motif GPxxxxxxxCSxSL around the active site Cys, and two His residues downstream from the active site Cys, in motifs characteristic of these enzymes. The KR domain contains the NAD(P)H-binding site GGxGxLG.

[00113] The PKS part of AlbI is linked by the PCP1 domain to the four apparent nonribosomal peptide synthase modules designated NRPS-1, NRPS-2, NRPS-3 and NRPS-4 (Figure 2). NRPS-1, NRPS-2 and NRPS-3 modules display the ordered condensation, adenylation (A) and PCP domains typical of such enzymes (Marahiel *et al.*, 1997), and NRPS-4 consists of an extra C domain which may correspond to an incomplete NRPS module. Known conserved sequences, characteristic of the domains commonly found in peptide synthases (Marahiel *et al.*, 1997), were compared to those from NRPS-1, NRPS-2, NRPS-3 and NRPS-4 (Tables 5, 6 and 7). Sequences characteristic of C, A, or PCP domains are conserved in these four NRPS, except in A domain of NRPS-2 module, suggesting that this latter A domain may be not functional. Comparison of the four NRPS modules among themselves revealed that NRPS-2, NRPS-3 and NRPS-4 modules were 30.7%, 94.4% and 47.5% similar to NRPS-1 module, respectively. Comparison with XabB revealed NRPS-2 and NRPS-3 modules were not present in XabB which contains only NRPS-1 and NRPS-4 modules (Figure 2). The dotted box in Figure 2 corresponds to the apparent deletion of the NRPS-2 and NRPS-3 modules in XabB as compared to AlbI. Apart their very high similarity with XabB, Alb I NRPS modules exhibited the highest degree of overall similarity with non-ribosomal peptide synthases NosA and NosC from *Nostoc* sp.

[00114] *albIV* potentially encodes a protein of 941 aa (AlbIV) with a predicted size of 104.8 kDa. AlbIV is similar to several non-ribosomal peptide synthases such as the BA3 peptide synthase involved in bacitracin biosynthesis in *Bacillus licheniformis* (Table 4). AlbIV forms one NRPS module designated NRPS-5 that contains only an A domain and a PCP domain (Figure 2). Sequences characteristic of the domains A and PCP commonly found in peptide synthases (Marahiel *et al.*, 1997) are conserved in AlbIV (Tables 6 and 7). However the A domain present in AlbIV differs from A domains commonly found in peptide synthases: conserved sequences corresponding to cores A8 and A9 in AlbIV are separated by a very long peptide sequence of 390 amino-acids. This additional peptide sequence exhibits a significative similarity with the protein WbpG of 377 amino acids involved in the biosynthesis of a lipopolysaccharide in *Pseudomonas aeruginosa* (Table 4).

[00115] *albVII* potentially encodes a protein of 765 aa (AlbVII) with a predicted size of 83.0 kDa similar to the 4-hydroxybenzoate-CoA ligase from several bacteria and the closest protein (HbaA) was from *Rhodopseudomonas palustris* (Table 4). High similarity between AlbVII and HbaA suggests that AlbVII is a 4-hydroxybenzoate-CoA ligase and constitutes a fourth PKS module designed PKS-4. The size of HbaA is smaller (539 aa) and the similarity between the two proteins starts only at the residue 277 of AlbVII and at the residue 28 of HbaA.

Comparison of AlbVII sequence located upstream from residue 277 produced no significant alignment. AlbVII, like 4-hydroxybenzoate-CoA ligases, contains some conserved sequences characteristic of the A domain commonly found in peptide synthases (Table 6).

[00116] *albIX* encodes a protein of 1959 aa (AlbIX) with a predicted size of 218.4 kDa similar to non-ribosomal peptide synthases. Known conserved sequences, characteristic of the domains commonly found in peptide synthases (Marahiel *et al.*, 1997), were compared with those from AlbIX which forms two NRPS modules designated NRPS-6 and NRPS-7 (Tables 5, 6 and 7). NRPS-6 contains only one A and one PCP domain. NRPS-7 contains the three domains characteristic of NRPS modules (A-C-PCP) followed by a TE domain (Figure 2). Apart their very high similarity with XabB, NRPS-6 and NRPS-7 modules exhibited the highest degree of overall similarity and identity with non-ribosomal peptide synthases DhbF from *B. subtilis* and NosA from *Nostoc* sp. (Table 4).

EXAMPLE 9 – The alb carbamoyl transferase gene

[00117] *albXV* potentially encodes a protein of 584 aa with a predicted size of 65.2 kDa. This protein, AlbXV, is similar to BlmD, a carbamoyl transferase involved in bleomycin biosynthesis in *Streptomyces vertillus* (Du *et al.*, 2000), and to a probable carbamoyl transferase potentially expressed in *P. aeruginosa* (Table 4). High similarity of AlbXV with these proteins suggests that AlbXV is a carbamoyl transferase.

EXAMPLE 10 – The alb esterase genes

[00118] *albXI* potentially encodes a protein of 315 aa with a predicted size of 35.9 kDa. This protein, AlbXI, exhibits low similarity to SyrC, a putative thioesterase involved in syringomycin biosynthesis by *Pseudomonas syringae* (Zhang *et al.*, 1995), and to a potential hydrolase encoded by *Streptomyces coelicolor* (Table 4). Precise function of SyrC remains unknown but SyrC is similar to a number of thioesterases, including fatty acid thioesterases, haloperoxidases, and acyltransferases that contain a characteristic GxCxG motif. The corresponding SyrC domain GICAG is conserved in AlbXI which contains the sequence GWCQA, except that A replaces the last G, suggesting that AlbXI may be an esterase despite its low overall similarity with SyrC.

[00119] *albXIII* potentially encodes a protein of 317 aa with a predicted size of 34.5 kDa. This protein, AlbXIII, is similar to hypothetical proteins with unknown function from several bacteria including *Caulobacter crescentus* (Table 4). AlbXIII and these hypothetical proteins contain a GxSxG motif characteristic of serine esterases and thioesterases, the corresponding sequence in AlbXIII being GHSVG. In addition, AlbXIII presents a similarity with

the 2-acetyl-1alkylglycerophosphocholine esterase which hydrolyzes the platelet-activating factor in *Canis familiaris* (Table 4), suggesting that AlbXIII is an esterase.

EXAMPLE 11 – The alb methyltransferase genes

[00120] *albII* potentially encodes a protein of 343 aa (AlbII) with a predicted size of 37.7 kDa. *albII* is 100% identical to the *xabC* cistron, previously described as encoding an *O*-methyltransferase downstream *xabB* (Huang *et al.*, 2000a). This conclusion is based on the similarity of XabC with a family of methyltransferases that utilize S-adenosyl-L-methionine (SAM) as a co-substrate for *O*-methylation including TcmO protein from *Streptomyces glaucescens* (Huang *et al.*, 2000a). AlbII contains three highly conserved motifs of SAM-dependent methyltransferases, including the motif I involved in SAM binding (Figure 3). In the Figure, identical or similar amino acids (A=G; D=E; I=L=V) are shown in bold. Numbers indicate the position of the amino acid from the N-terminus of the protein. Abbreviations used in the Figure are: Sgl-TcmO and Sgl-TcmN, multifunctional cyclase-hydratase-3-*O*-Mtase and tetracenomycin polyketide synthesis 8-*O*-Mtase of *Streptomyces glaucescens*, respectively (accession number: M80674); Smy-MdmC, midecamycin-*O*-Mtase of *Streptomyces mycarofaciens* (accession number: M93958); Mxa-SafC, Saframycin *O*-Mtase of *Myxococcus xanthus* (accession number: U24657); Ser-EryG, erythromycin biosynthesis *O*-Mtase of *Saccharopolyspora erythraea* (accession number: S18533); Spe-DauK, carminomycin 4-O-Mtase of *Streptomyces peucetius* (accession number: L13453); Sal-DmpM, *O*-demethylpuromycin-*O*-Mtase of *Streptomyces alboniger* (accession number: M74560); Shy-RapM, rapamycin *O*-Mtase of *Streptomyces hygroscopicus* (accession number: X86780); Sav-AveD, avermectin B 5-*O*-Mtase of *Streptomyces avermitilis* (accession number: G5921167), Sar-Cmet, mithramycin *C*-methyltransferase of *Streptomyces argillaceus* (accession number: AF077869); AlbII, putative albicidin biosynthesis *C*-Methyltransferase of *Xanthomonas albilineans* (SEQ ID No. 27); identical to XabC, accession number: AF239749).

[00121] Comparison of AlbII with the Genbank database revealed that AlbII, besides 100% identity to XabC, exhibited the highest degree of overall identity with MtmMII, a *C*-methyltransferase from *Streptomyces argillaceus* (Table 4) involved in *C*-methylation of the polyketide chain for mithramycin biosynthesis, suggesting that AlbII is a *C*-methyltransferase. XabC was not compared by Birch and co-workers with MtmMII (Huang *et al.*, 2000a) because the MtmMII sequence was not available until recently in the Genbank database. The three highly conserved motifs in SAM methyltransfererases are also present in MtmMII (Figure 3), suggesting that AlbII is a *C*-methyltransferase SAM-dependent.

[00122] *albVI* potentially encodes a protein of 286 aa (AlbVI) with a predicted size of 32.1 kDa similar to several hypothetical protein from *Mycobacterium tuberculosis*

(Genbank accessions No. AAK46042, AAK48238, AAK44517, AAK46218) and from *S. coelicolor* (Genbank accession No. CAC03631). AlbVI is also similar to the tetracenomycine C synthesis protein (TcmP) of *Pasteurella multocida* (Table 4). Four highly conserved motifs in TcmP and other *O*-methyltransferases are also present in AlbVI (Figure 4), suggesting that AlbVI is an *O*-methyltransferase. In the Figure, identical or similar aa (A=G; D=E; I=L=V; K=R) are shown in bold. Numbers indicate the position of aa from the N-terminus of the protein. Abbreviations used in the Figure are: Sgl-tcmP, tetracenomycin C synthesis protein of *Streptomyces glaucescens* (accession number: C47127); Sme-PKS, putative polyketide synthase of *Sinorhizobium meliloti* (accession number: AAK65734); Pmu-tcmP: tetracenomycin C synthesis protein of *Pasteurella multocida* (accession number: AAK03406); Mtu-Omt: putative *O*-methyltransferase of *Mycobacterium tuberculosis* (accession number: AAK45444); Mlo-Hp: hypothetical protein containing similarity to *O*-methyltransferase of *Mesorhizobium loti* (accession number: BAB50127); Mtu-Hp1: hypothetical protein of *Mycobacterium tuberculosis* (accession number: AAK46042); Mtu-Hp2: hypothetical protein of *Mycobacterium tuberculosis* (accession number: AAK48238); Mtu-Hp3: hypothetical protein of *Mycobacterium tuberculosis* (accession number: AAK44517); AAK46218); Sco-Hp: hypothetical protein of *Streptomyces coelicolor* (accession number: CAC03631); AlbVI, putative albicidin biosynthesis *O*-Methyltransferase of *Xanthomonas albilineans* (this study). The three highly conserved motifs in SAM methyltransfererases are not present in AlbVI, indicating that SAM is not a co-substrate of AlbVI.

EXAMPLE 12 – The alb derived-benzoate products biosynthesis genes

[00123] alb*XVII* potentially encodes a protein of 716 aa with a predicted size of 79.8 kDa. This protein, AlbXVII, is very similar to the para-aminobenzoate (PABA) synthase from *Streptomyces griseus* (Table 4). This enzyme is required for the production of the antibiotic candicidin (Criado *et al.*, 1993).

[00124] *albXVIII* potentially encodes a protein of 137 aa with a predicted size of 15.0 kDa. This protein, AlbXVIII, is similar to the 4-amino-4-deoxychorismate lyase (ADCL) from *P. aeruginosa* (Table 4). The function of ADCL is to convert 4-amino-4-deoxychorismate into PABA and pyruvate. The length of AlbXVIII is smaller (Table 4) than the length of ADCL and the similarity of AlbXVIII with this protein starts only at residue 161. *albXVIII* is preceded by a small ORF encoding a sequence of 59 amino acids similar to the first 42 amino acids of ADCL from *P. aeruginosa*. These data suggest that *albXVIII* is probably a truncated form of *albXVIII* and probably not functional. *albXVIII* may, therefore, not be involved in albicidin biosynthesis. The region between *albXVII* and *albXVIII* was amplified by PCR from total DNA of *X. albilineans* Xa23R1 strain using primers ORFW (5'gcgagaggacaagctgctgc3') SEQ ID No. 53 and ORFY (5'cgttgaggatgcagcgctcg3') SEQ ID No. 54 and was sequenced. Resulting sequence data

showed that the sequence of the PCR fragment was 100% identical to the sequence of pALB540, indicating that the recombination of *albXVIII* did not occur during cloning of the genomic fragment in pALB540.

[00125] *albXX* potentially encodes a protein of 202 aa with a predicted size of 22.6 kDa. This protein AlbXX is similar to the 4-hydroxybenzoate synthase potentially involved in ubiquinone biosynthesis by *Escherichia coli* (Siebert *et al.*, 1992).

EXAMPLE 13 – The alb regulatory genes

[00126] *albIII* potentially encodes a protein of 167 amino acids with a predicted size of 17.8 kDa that is similar to the transcription factors ComA of different bacteria such as *E. coli* and *B. licheniformis* (Table 4). ComA transcription factors appear to be involved in regulation of antibiotic production in bacteria. In *E. coli*, a gene similar to *comA* is present in the enterobactin biosynthesis gene cluster (Liu *et al.*, 1989). In *B. subtilis*, ComAB was described as a probable positive activator of lichenysin synthetase transcription (Yakimov *et al.*, 1998) and a gene similar to *comA* was shown to be essential for bacilysin biosynthesis (Yazgan *et al.*, 2001). These data suggest that AlbIII regulates transcription of genes involved in albicidin biosynthesis.

[00127] *albVIII* potentially encodes a protein of 330 aa with a predicted size of 37.7 kDa. This protein, AlbVIII, is very similar to the SyrP like protein from *S. verticillus* and to SyrP protein from *P. syringae* (Table 4). SyrP participates in a phosphorylation cascade controlling syringomycin synthesis (Zhang *et al.*, 1997) and the *syrP* like gene was described in the *S. verticillus* bleomycin biosynthetic gene cluster (Du *et al.*, 2000). These data suggest that AlbVIII regulates albicidin biosynthesis in *X. albilineans*.

EXAMPLE 14 – The alb resistance genes

[00128] *albXIV* potentially encodes a protein of 496 aa with a predicted size of 52.7 kDa. This protein, AlbXIV, is 100% identical to AlbF isolated from *X. albilineans* strain Xa13 (GenBank Accession AF403709; direct submission by Bostock and Birch and described as "a putative albicidin efflux pump which confers resistance to albicidin in *E. coli*"). AlbXIV and AlbF are closely related to a family of transmembane transporters involved in antibiotic export and antibiotic resistance in many antibiotic-producing organisms. AlbXIV and AlbF exhibited the highest degree of overall identity with the putative transmembrane efflux protein from *S. coelicolor* (Table 4). These data suggest that AlbXIV and AlbF may be involved in albicidin resistance by transporting the toxin out of the bacterial cells that produce it. Alternatively, AlbXIV and AlbF may simply play a role in antibiotic secretion and/or plant pathogenesis to effect the transport of albicidin outside of producing cells.

[00129] *albXIX* potentially encodes a protein of 200 aa with a predicted size of 22.8 kDa. This protein, AlbXIX, is similar to the McbG protein from *E. coli* (Table 4). In *Enterobacteriae*, the McbG protein, together with two other proteins (McbE and McbF), was shown to cause immunity to the peptide antibiotic microcin B17 which inhibits DNA replication by induction of the SOS repair system (Garrido *et al.*, 1988). McbE and McbF proteins serve as a pump for the export of the active antibiotic from the cytoplasm, whereas a McbG alone also provides some protection: a well-characterized deficient-immunity phenotype is exhibited by microcin B17-producing cells in the absence of the immunity gene mcbG (Garrido *et al.*, 1988). The significant similarity between AlbXIX and McbG, together with the fact that albicidin also blocks DNA replication (Birch and Patil, 1985a) suggests that AlbXIX confers immunity against albicidin in *X. albilineans*.

EXAMPLE 15 – Transposition proteins

[00130] *albV* is 100% identical to the *thp* gene described in a divergent position upstream from *xabB* (Huang *et al.*, 2000a). The *thp* gene potentially encodes a protein of 239 aa displaying significant similarity to the IS21-like transposition helper proteins. In *X. albilineans* strain LS155 from Australia, insertional mutagenesis of *thp* blocked albicidin production, but *trans*-complementation failed, indicating the involvement in albicidin production of a downstream gene in the *thp* operon (Huang *et al.*, 2000a).

[00131] *albXVI* potentially encodes a protein of 88 aa with a predicted size of 9.8 kDa similar to the transposases from several bacteria such as *Xanthomonas axonopodis* or *Desulfovibrio vulgaris* (Table 4).

[00132] The presence of transposition proteins in the XALB1 cluster is probably a remnant from a past transposition event that may have contributed to the development of the albicidin XALB1 cluster.

EXAMPLE 16 – Unknown functions

[00133] *AlbX* potentially encodes a protein of 83 aa with a predicted size of 9.4 kDa. This protein, AlbX, is similar to an hypothetical protein from *P. aeruginosa* and to the MbtH protein from *Mycobacterium tuberculosis*. MbtH is a protein with unknown function found in the mycobactin gene cluster (Quadri *et al.*, 1998). A MbtH-like protein with unknown function was also described in the bleomycin biosynthetic gene cluster of *S. verticillus* (Du *et al.*, 2000). These data suggest that AlbX is involved in albicidin biosynthesis but its function remains unknown.

[00134] *albXII* potentially encodes a protein of 451 aa with a predicted size of 51.6 kDa. This protein, AlbXII, is very similar to a protein of 55 kDa encoded by the *boxB* gene in *Azoarcus evansii* (Table 4). This protein is a component of a multicomponent enzyme system

involved in the hydroxylation of benzoyl CoA, a step of aerobic benzoate metabolism in *Azoarcus evansii*, but its function remains unknown (Mohamed *et al.*, 2001).

EXAMPLE 17 – Prediction of amino acid specificity of Alb NRPS modules

[00135] In NRPSs, specificity is mainly controlled by A domains which select and load a particular amino-, hydroxy- or carboxy-acid unit (Marahiel *et al.*, 1997). The substrate-binding pocket of the phenylalanine adenylation (A) domain of the gramicidin S synthetase (GrsA) from *Brevibacillus brevis* was recently identified by crystal structure analysis as a stretch of about 100 amino acid residues between highly conserved motifs A4 and A5 (Conti *et al.*, 1997). Based on sequence analysis of known A domains, in relation to the crystal structure of the GrsA (Phe)substrate binding pocket, similar models have been published to predict the amino acid substrate which is recognized by an unknown NRPS A domain (Challis *et al.*, 2000; Stachelhaus *et al.*, 1999). These models postulate specificity-conferring codes for A domains of NRPS consisting of critical amino acid residues putatively involved in substrate specificity. The model proposed by Marahiel and co-workers (Stachelhaus *et al.*, 1999) defined a signature sequence consisting of ten amino acids lining with the ten residues of the phenylalanine-specific binding pocket located at positions 235, 236, 239, 278, 299, 301, 322, 330, 331 and 517 in the GsrA (Phe) sequence (accession number: P14687). The model proposed by Townsend and co-workers (Challis *et al.*, 2000) uses only the first eight of these critical residues.

[00136] Preliminary specificity assignments of albicidin synthase AlbI, AlbIV, AlbVII and AlbIX NRPS modules were made by comparison of complete sequences between conserved motifs A4 and A5 with sequences in the Genbank database. The corresponding sequence of the AlbIV NRPS-5 module is most related to domain 5 of bacitracin synthase 3 (BA3) from *B. licheniformis* that was suggested to activate Asn (Konz *et al.*, 1997). Corresponding sequences of AlbI and AlbIX NRPS-1, NRPS-3, NRPS-6 and NRPS-7 modules, apart from their very high similarity with XabB, exhibited the highest degree of overall identity (39%) with the Blm NRPS2 module of the biosynthetic gene cluster for bleomycin from *S. verticillus* that specifies for β-Alanine (Du *et al.*, 2000). The corresponding sequence of AlbVII PKS-4 produced the highest significant alignment with acetate-CoA ligase from *Sulfolobus solfataricus* (Genbank accession number: AAK41550), aryl-CoA ligase from *Comamonas testosteroni* (Genbank accession number: AAC38458) and 4-hydroxybenzoate-CoA ligase from *R. palustris*. The sequence between motifs A4 and A5 of the AlbI NRPS-2 could not be significantly aligned with any sequence present in the Genbank database. Comparison of this sequence with the corresponding sequence of GrsA (Phe) revealed that parts of the putative core and structural "anchor" sequences of AlbI NRPS-2 are deleted (Figure 5), suggesting that the AlbI NRPS-2 substrate binding pocket is not functional. In the Figure, amino acids of the six Alb NRPSs and of Alb PKS-4 that are identical or similar to GrsA or Blm sequences (A=G; D=E;

I=L=V; R=K) are shown in bold. Amino acids underlined in the GsrA sequence correspond to the phenylalanine-specific binding pocket. The positions of these amino acids in the GrsA primary sequence are indicated at the top of the figure. Amino acids underlined in the other sequences correspond to putative constituents of binding pockets, aligned with the seven residues of the phenylalanine- specific binding pocket of GrsA. Shaded amino-acids correspond to the putative core sequences and structural "anchors" based on comparison with the GrsA binding-pocket structure.

[00137] Alignment of the primary sequence between conserved motifs A4 and A5 of the AlbI, AlbIV, AlbVII and AlbIX NRPS-1, NRPS-3, NRPS-5, NRPS-6, NRPS-7 and PKS-4 modules with the corresponding sequence of GrsA (Phe) (Figure 5) revealed the putative constituents of binding pockets that constitute the codes as defined by Marahiel and co-workers (Stachelhaus *et al.*, 1999). These codes were compared with those of proteins most related to the sequence between the A4 and A5 motifs (Table 8) and were analyzed with the model proposed by Townsend and co-workers (Challis *et al.*, 2000, jhunix.hcf.jhu.edu/ ~ravel/nrps//). Using these codes, we were able to predict the Asparagine specificity of the AlbIV NRPS-5 module. The AlbIV NRPS-5 signature is 100% identical to BacC-M5 (Asn) and TyrC-M1 (Asn) codes identified in bacitracin synthetase 3 from *B. licheniformis* and in tyrocidine synthetase 3 from *B. brevis* (Table 8). The AlbIV NRPS-5 signature is also identical to the Asn code defined by Marahiel and co-workers (1997), except that I is replaced by L at position 299 (Table 8). The AlbI and AlbIX NRPS-1, 3, 6 and 7 signatures did not match any of those defined by Marahiel and co-workers (1997). Similarly, convincing predictions using the model proposed by Townsend and co-workers were not obtained either (Challis *et al.*, 2000, jhunix.hcf.jhu.edu/~ravel/nrps//). The AlbI and AlbIX NRPS-1, 3, 6 and 7 signatures diverged from all NRPS signatures previously described, except from the XabB signature that is identical to the AlbI NRPS-1 and 3 signatures. The signature most closely related to AlbI NRPS-1 and 3 specify Pro and the signature most closely related to AlbIX NRPS-6 and 7 specify Ser, but the degree of similarity in both cases is very weak (Table 8). The PKS-4 signature is similar to the AlbI NRPS-1 and NRPS-3 signatures at positions 235, 299 and 301.

[00138] Analysis of alignment of the primary sequence between conserved motifs A4 and A5 of the AlbI and AlbIX NRPS-1, NRPS-3, NRPS-6 and NRPS-7 modules with the corresponding sequences of the bleomycin synthase (Blm) NRPS2 (β-Ala) and gramicidin S synthetase (GrsA) modules (Figure 5) revealed that (i) sequences of AlbI NRPS-1 and AlbI NRPS-3 differ only at the level of two residues that are not involved in substrate binding, (ii) sequences of AlbIX NRPS-6 and AlbIX NRPS-7 are 100% identical, (iii) sequences of AlbI NRPS-1 and AlbI NRPS-3 are very similar to sequences of AlbIX NRPS-6 and AlbIX NRPS-7 but differ at the level of five putative constituents of binding pocket, (iv) AlbI and AlbIX NRPS residues, which are similar to residues of Blm NRPS2 (β-Ala) or GrsA (Phe), are essentially

located at the level of the putative core sequences and structural "anchor", and differ at the level of putative constituents of the binding pocket.

[00139] Binding-pocket constituents forming the NRPS signatures have been classified into three subgroups according to their variability among 160 specificity-conferring signature sequences (Stachelhaus *et al.*, 1999): (i) invariant residues Asp235 and Lys517 that mediate key interactions with the α-amino and α-carboxylate group of the substrate, respectively; (ii) moderately variant residues in positions 236, 301 and 330 which correspond to aliphatic amino acids and which may modulate the catalytic activity and fine-tune the specificity of the corresponding domains; (iii) highly variant residues in positions 239, 278, 299, 322 and 331 which may facilitate substrate specificity. AlbI and AlbIX NRPS-1, 3, 6 and 7 signatures are not totally in accordance with this classification. Invariant residue Lys517 is conserved in the four NRPS signatures, indicating the presence of an α-carboxylate group in the corresponding substrates. The Asp235Ala alteration is not consistent with an α-amino acid substrate. Birch and co-workers (Huang *et al.*, 2001) assumed that the initial alanine residue in the XabB signature was consistent with a nonproteinogenic hydroxy acid substrate by analogy with the initial glycine in the signature of the hydroxyisovaleric-acid (HVCL) loading domain of enniatin synthetase. The presence of an initial Alanine in the AlbVII PKS-4 signature (Figure 8) and in several 4-hydroxybenzoate-CoA ligase codes may confirm this hypothesis. However, the HVCL loading domain of enniatin synthetase (Table 8) and AlbVII PKS-4 are not preceded by a C domain and are not followed by a PCP domain, in contrast to the AlbI and AlbIX NRPS-1, 3, 6 and 7 modules. An Asp235Val alteration was recently described in the β-Ala specificity-conferring code (Du *et al.*, 2000, Table 8), suggesting that the substrate of AlbI and AlbIX NRPS-1, 3, 6 and 7 modules may be different from α-amino acids but may contain an amino group. Residue 236 is an aliphatic residue (Val or Ile) in all AlbI and AlbIX NRPS-1, 3, 6 and 7 signatures. Residue 301 is an aliphatic residue (Ala) in the AlbI NRPS-1 and 3 codes, but it is a Ser in the AlbIX NRPS-6 and 7 signatures. Residue 330 is not an aliphatic residue in the four NRPS signatures but an Asp. Similar alterations are present in the β-Ala code: residue 236 is an Asp, residue 301 is a Ser and residue 330 is an aliphatic amino acid. Concerning highly variable residues, AlbI NRPS-1 and 3 signatures differ from AlbIX NRPS-6 and 7 signatures at residue positions 299, 322 and 331, confirming that both types of NRPS modules specify different substrates.

[00140] **Table 8 : Comparison of signature sequences, as defined by Marahiel and co-workers** (Stachelhaus *et al.*, 1999), derived from sequences between the A4 and A5 domains of the AlbI, AlbIV, and AlbIX NRPS modules with those of Tyr-M1 (Pro) (Tyrocidine synthetase 2 module 1, accession number: AAC45929), VirS (Pro) (Virginiamycin S synthetase, accession number : CAA72310), HVCL (hydroxyisovaleric acid-CoA ligase, ACL1 enniatin synthetase, accession number: S39842), EntF-M1 (Ser) (Enterobactin synthase, accession number: AAA92015), β-Ala code (β-Ala selectivity-conferring code defined by Du *et al.* , 2000),

BacC-M5 (Asn) (Bacitracin synthetase 3, accession number: AAC06348), TyrC-M1 (Asn) (Tyrocidine synthetase 3, accession number: AAC45930) and Asn code (Asn selectivity-conferring code defined by Marahiel and co-workers (Stachelhaus *et al.*, 1999). Amino acids of AlbI NRPS-1 and NRPS-3 signatures identical or similar to TyrB-M1 (Pro), VirS (Pro) and HVCL signatures (A=G; D=E; I=L=V; R=K) are shown in bold. Amino acids of AlbIX NRPS-6 and NRPS-7 signatures identical or similar to Vir (Pro) and Blm (ß-Ala) signatures (A=G; D=E; I=L=V; R=K) are shown in bold. Variability: 0 indicates invariant residues, +/- moderately variant residues and ++ highly variant residues.

EXAMPLE 18 – Identification of putative promoters and putative terminators in XALB1

[00141] Putative rho independent terminators were identified downstream from *albIV* and *albXVI* using the Terminator program (Brendel and Trifonov, 1984), run with the Wisconsin Package™ GCG software (Figure 6). In the Figure, dashes indicate palindromic sequences. Symbols used in the Figure are: P, Primary structure value of putative terminator (minimum threshold value of 3.5 represents 95 percent of known, factor-independent, prokaryotic terminators); S, Secondary structure value of putative terminator. The presence of these terminators confirmed the proposed genetic organization of operons 1 and 3. A rho-independent terminator was identified in the intergenic region between *albXVII* and *albXVIII*, suggesting that the group of genes initially supposed to be organized in operon 4 may be in fact organized in two operons, operon 4 formed by *albXVII* and operon 5 by *albXVIII* B *albXX*. No putative rho independent terminator was found downstream from *albIX* and from *albXX*.

[00142] The 236 bp region between *albI* (operon 1) and *albV* (operon 2) is 100% identical to the sequence between *xabB* and *thp* genes that is assumed to contain a bidirectional promoter (Huang *et al.*, 2000a and 2001), suggesting that transcription of operon 1 and 2 is regulated by the same bidirectional promoter region (Huang *et al.*, 2001).

[00143] The 412 bp region comprised between *albX* (operon 3) and *albXVII* (operon 4) also contains a putative bidirectional promoter (Figure 7). In the Figure, the sequence of putative promoters are underlined, and putative ATG or TTG start codons are in bold. The closest matches (TTGACA-18x-TATAGT) to the consensus -35 (TTGACA) and -10 (TATAAT) sequences for *E. coli* σ^{70} promoters occurs 61 bp upstream from *albX* (operon 3). The closest matches (TTCAGA-19x-TATACA) to the consensus sequences for *E. coli* σ^{70} promoters occur 320 bp upstream from *albXVII* (operon 4). The region between *albXVII* and *albXVIII* lacks any apparent *E. coli* σ^{70} promoter. However, the sequence immediately upstream from *albXIX*, corresponding to the coding sequence of *albXVIII*, potentially contains an unidirectional promoter (Figure 7). The closest match (TTGCTC-19x-TATATT) to the consensus sequences for *E. coli* σ

^{70} promoters occurs 33bp upstream from *albXIX*. The presence of a terminator downstream from *albXVII* and of a promoter upstream from *albXIX* suggests that *albXVIII* is not transcribed and that *albXIX* and *albXX* form operon 5.

EXAMPLE 19 – Cloning of the XALB2 gene cluster

[00144] The 6 kb *Eco*R I fragment carrying Tn*5* and flanking sequence from strain AM37 was cloned in pBR325 and the obtained plasmid was designated pAM37 (Table 1). A 1.1 kb *Hin*d III-*Hin*d III DNA fragment from pAM37, named PR37 (Table 1), was labeled with ^{32}P and used to probe the 845 clones from the genomic library of *X. albilineans* strain Xa23R1, previously described (Rott et al., 1996). Eight new cosmids hybridized to this probe and restored albicidin production in mutant AM37. One of these cosmid, pALB389, carrying an insert of about 37 kb (Table 1), was used for complementation studies of the five mutants not complemented by pALB540 and pALB571. Cosmid pALB389 complemented mutants AM10 and AM37. Mutant AM10 was initially thought to be complemented by pALB639 (Rott et al., 1996). However, further complementation studies showed that mutant AM10 was not complemented by pALB639 and that only three mutants (AM12, AM13 and AM36) were complemented by pALB639 containing the third genomic region XALB3 involved in albicidin production. A 3 kb *Eco*RI- *Eco*RI DNA fragment from pALB389 that hybridized with probe PR37 was sub-cloned in pUFR043 (Table 1). The resulting plasmid pAC389.1 complemented mutants AM10 and AM37, confirming that the second region involved in albicidin production, XALB2, was present in the 3 kb insert of pAC389.1.

EXAMPLE 20 – Cloning of the XALB3 gene cluster

[00145] Cosmid pALB639, carrying an insert of 36 kb (Rott *et al.*, 1996; Table 1) was used as a probe to compare the *Eco*RI restriction profiles of *X. albilineans* strain Xa23R1 with those of mutants AM12, AM13 and AM36 which were supposed to be mutated in the XALB3 gene cluster. An 11 kb band which was found in strain Xa23R1 but not in the three mutants was supposed to contain the XALB3 gene cluster. A 9.7 kb *Eco*RI DNA fragment purified from cosmid pALB639 also used as a probe in Southern blot analyse revealed the same 11 kb band. This 9.7 kb *Eco*RI DNA fragment was sub-cloned in pUFR043 (Table 1) and the resulting plasmid pAlb639A complemented mutants AM12, AM13 and AM36. The third region involved in albicidin production, XALB3, was therefore present in the 9.7 kb insert of pAlb639A.

EXAMPLE 21 – Sequencing of the Tn*5* insertional site of tox[-] mutants located in XALB2 and XALB3 and sequencing of the genomic regions XALB2 and XALB3

[00146] In Figure 8, E, H, Sa and S indicate restriction endonuclease cut sites for *Eco*RI, *Hin*dIII, *Sal*I and *Sau*3AI, respectively. The DNA inserts carried by plasmids pAC389.1, pALB639A or pEV639 are represented by the bars at the top of the respective figures. Positions of the Tn*5* insertional sites of mutants AM10, AM12, AM36 and AM37 were determined by sequencing and are indicated by vertical arrows. The DNA region corresponding to the Tn*5* flanking regions in pAM10, pAM12.1, pAM36.2 and pAM37 and in the PR37 DNA fragment are represented by the bars at the bottom of the respective figures. The location and direction of *albXXI* and *albXXII* are indicated by thick black arrows. The location of other orfs in XALB2 similar to those described by Huang *et al.* (2000b) are indicated by thick white arrows.

[00147] The 7 kb *Eco*R I fragment carrying Tn*5* and flanking sequence from strain AM10 was cloned in pBluescript II KS (+), and the obtained plasmid was designated pAM10 (Table 1). The sequences between *Eco*RI sites and the Tn*5* insertional site of mutants AM10 and AM37 were sequenced from the resulting plasmids pAM10 and pAM37, respectively. The complete double-strand nucleotide sequence of the 2,986 bp *Eco*R I B *Eco*R I insert of pAC389.1 was determined from sequencing results of plasmids pAC389.1, pAM10 and pAM37 (Figure 8). The Tn*5* insertional sites of mutants AM10 and AM37 were sequenced from plasmids pAM10 and pAM37 (Table 1), respectively, using the sequencing primer GUSN (5'tgcccacaggccgtcgagt3') that annealed 135 bp downstream from the insertional sequence IS50L of Tn*5*-*gusA*. The Tn*5* insertional site of AM10 and AM37 was located at position 2107 and 1882, respectively.

[00148] The *Eco*RI fragments carrying Tn*5* and the flanking sequences from mutants AM12 and AM36 were cloned in pBR325 (Rott *et al.*, 1996; Table1). The sequences between *Eco*RI site and the Tn*5* insertional site of mutants AM12 and AM36 were sequenced from the resulting plasmids pAM12.1 and pAM36.2, respectively. The complete double-strand nucleotide sequence of the 9,673 bp *Eco*R I B *Sau*3A I insert of pALB639A was determined from the sequencing results of plasmids pAM12.1, pAM36.2 and pALB639A (Figure 8). The Tn*5* insertional site of mutants AM12 and AM36 was sequenced from plasmids pAM12.1, pAM36.2 using the sequencing primer GUSN (5'tgcccacaggccgtcgagt3') that annealed 135 bp downstream from the insertional sequence IS50L of Tn*5*-*gus*A. The Tn*5* insertional site of AM12 and AM36 was located at position 6500 and 7232, respectively (Figure 8).

EXAMPLE 22 – Homology analysis and genetic organization of XALB2 (Figure 8).

[00149] The sequence of 2986 bp containing XALB2 is 99.4% identical to the sequence of 2989 bp containing *xabA* described in *X. albilineans* strain LS155 from Australia (Huang et al., 2000b; accession number AF191324). The Tn*5* insertional site of mutant LS156 described in *xabA* is 15 bp upstream from the insertional site of AM37. The orf disrupted in AM37 and AM10, designed albXXI, is identical to *xabA* except a C which replaces a T at position 1642. albXXI potentially encodes a protein of 278 aa with a predicted size of 29.3 kDa which is 100% identical to the potential product of xabA, described as a phosphopantetheinyl transferase (Huang et al., 2000b). Region XALB2 contains three additional orfs (orf1, orf2, and orf3) similar to those described by Huang et al., (2000b; orf, rsp6 and aspT). orf2 and orf3 are 100% identical to rsp6 and aspT respectively, and orf1 is similar to but smaller than orf. There are no close matches to the *E. coli* $\gamma 70$ promoter B10 (TATAAT) and B35 (TTGACA) consensus sequence, and no putative RBS site upstream from the putative start codon ATG of albXXI. The putative factor-independent transcription site described at 42 bp downstream from the TGA stop codon of xabA (Huang et al., 2000b) is also present at the same position downstream from albXXI.

EXAMPLE 23 – Homology analysis and genetic organization of XALB3 (Figure 8).

[00150] The orf disrupted in mutants AM12 and AM36 was located between nucleotide 6090 (ATG) and 8009 (TAA) and was designed *albXXII*. The first ATG at position 6090 is not preceded by a putative ribosome binding sequence, suggesting that the start codon is the ATG at position 6105 which is preceded at position B9 by the putative ribosome binding site sequence GGAG. A putative rho independent terminator was identified at position 8082, 73 b downstream from *albXXII* (Figure 6). There are no close matches to *E. coli* σ^{70} promoter B10 (TATAAT) and B35 (TTGACA) consensus sequence upstream from the putative start codon. The *Sal*I DNA fragment corresponding to DNA sequence from nucleotide 5510 to nucleotide 8124, which contains the 595 bp upstream from the putative start codon, the orf *albXXII* and the putative rho independent terminator, was sub-cloned in pUFR043 in the opposite direction to LacZ (Table 1). The resulting plasmid pEV639 (Table 1) complemented mutants AM12, AM13 and AM36, confirming that (i) the third region involved in albicidin production, XALB3, was present in the insert of pEV639; (ii) *albXXII* is not transcribed as part of a larger operon; and (iii) the 595 bp upstream the putative start codon contain a promoter.

[00151] The potential product of *albXXII*, designated AlbXXII, is a protein of 634 aa with a predicted size of 71.5 kDa. This protein is very similar to the heat shock protein HtpG from *Pseudomonas aeruginosa* (identities = 82%) and from *Escherichia coli* (identities = 60%) (Table 4). The methionine encoded by the putative start codon at position 6105 was aligned with

the first aminoacid of the heat shock protein HtpG from *Pseudomonas aeruginosa*, confirming that *albXXII* initiates at position 6105.

Complementation of Tox⁻ mutants with the *albXXII* gene in fusion with LacZ

[00152] A 1,948 bp fragment corresponding to the entire 1,903 bp orf of *albXXII* and flanking nucleotides was PCR amplified from cosmid pALB639 with the forward primer 5'tttgaattcgcacctaccgatgagcgtgg3' and the reverse primer 5'tttggatccgtgcgtcactgcttacgccg3'. Convenient in frame-*Eco*RI and *Bam*HI restriction sites for further cloning were simultaneously introduced with forward and reverse PCR primers, respectively. The PCR fragment was cloned into pGEMT vector (Promega) and sequenced. Several clones of the resulting plasmid pGemT/albXXII were sequenced. Because several PCR derived point mutations were observed in all the sequenced clones, a 1,920 bp *Bgl*II – *Sal*I fragment from pEV639 (corresponding to the 1,809 5' terminal nucleotides of *albXXII* orf plus 111 bp downstream the stop codon) was cloned into a pGemT/albXXII clone between the *Bgl*II site located at position 94 of the *albXXII* orf and the *Sal*I site of the vector's multiple cloning site. The resulting plasmid pGemT/albXXIIbis contained an intact *albXXII* orf that was then subcloned as an *Eco*RI – *Sal*I fragment into pUFR043 to generate pEValbXXII. This construct of *albXXII* in fusion with LacZ was transferred by triparental conjugation into Xa23RI insertion mutants. pEValbXXII complemented mutants AM12, AM13 and AM36 (see table 9). These results confirmed that (i) the third region involved in albicidin production, XALB3, was present in the insert of pEValbXXII; and (ii) *albXXII* is not transcribed as a part of a larger operon.

Complementation of Tox⁻ mutants with the *htpG* gene from *E. coli*

[00153] A 2,343bp fragment corresponding to the *htpG* gene of *E. coli* plus 458 bp downstream the stop codon was PCR amplified from purified DH5α genomic DNA with forward primer 5'tttgaattccatgaaaggacaagaaactcgtgg3' and reverse primer 5'gcctgcggaatggtacgcgggaagccgtcc3'. A convenient in frame-*Eco*RI restriction site was introduced with the forward PCR primer. The PCR fragment was cloned using the pGEMT vector system (Promega). Three resulting clones potentially containing plasmid pGemT/HtpG were sequenced, and one clone containing the correct sequence was selected. The 2,343bp PCR insert was then subcloned as an *Eco*RI – *Sal*I fragment into pUFR043 to generate pEVHtpG, the *Sal*I site corresponding to the site of the vector's multiple cloning site. This HtpG gene, in fusion with the LacZ construct, was able to restore albicidin production after transfer by triparental conjugation into AM12, AM13 and AM36 Xa23RI mutants. This result is i/ further evidence of the involvement of a molecular chaperone HtpG in the biosynthesis of albicidin (table 9), ii/ the first report of the requirement of a molecular chaperone HtpG in NRPS and PKS metabolism.

EXAMPLE 24 - Heterologous production of albicidin in fast growing *Xanthomonas axonopodis* pv. Vesicatoria.

[00154] This example illustrates the construction of a heterologous expression system harboring the three XALB regions, its transfer into a fast growing host, *Xanthomonas axonopodis* pv. vesicatoria and the subsequent production of a potent toxin with an antibiotic activity similar to that of albicidin. This work is a milestone in the validation of the albicidin biosynthesis model because it gives experimental evidence that the entire biosynthetic machinery required for albicidin biosynthesis has been identified, cloned, sequenced and transferred into an heterologous host, driving the production of albicidin. Cosmid pALB571 which covers the complete sequences of operons 1 and 2 was used to transfer operons 1 and 2 (Figure 1). Operons 3 and 4 (from pALB540), XALB2 (from pAC389.1) and XALB3 (from pEV639) were subcloned into a single plasmid, pOp3-4/XALB2-3 (see below). Plasmid pOp3-4/XALB2-3 derived from shuttle vector pLAFR3 that carries one selective gene for resistance to tetracyclin and that belongs to incompatibility group IncP (Table 1). Cosmid pALB571 derived from shuttle vector pUFR043 that carries two selective genes for resistance to kanamycine and gentamycine and that belongs to incompatibility group IncW (Table 1).

Sub-cloning of operons 3 and 4 and XALB2 and XALB3 regions into a single plasmid (Figure 12).

[00155] A 2,787 bp *Bam*HI – *Pst*I fragment from pALB540, corresponding to a portion of operon 4, was subcloned into pBCKS(+), yielding pBC/Op4Δ (step 1). A *Xho*I site was introduced into this vector immediately upstream from the *Bfr*I site by directed mutagenesis. Mutagenesis was performed with primers XhoIAlb anticodant 5'cgccttaagcagctcgagtagactgcaatc3' and XhoIAlbcodant 5'gattgcagtctactcgagctgcttaaggcg3' and yielded plasmid pBC/Op4ΔXhoI (step 2). The 2,986 bp *Eco*RI fragment from pAC389.1 (containing XALB2) was then subcloned into pBC/Op4ΔXhoI, yielding pBC/Op4Δ/XALB2 (step 3). A 10,762 bp *Bfr*I fragment from pALB540 and containing complete operon 3 and the beginning of operon 4 was subcloned into pBC/Op4Δ/XALB2 yielding pBC/Op3-4/XALB2 (step 4). The 2,615 bp *Sal*I fragment from pEV639 (containing XALB3) was subcloned into pBKS, yielding pBKS/XALB3 (step 5). The *Sal*I site located on the *Kpn*I side of the polylinker was then destroyed and substituted by a *XhoI* restriction site by directed mutagenesis. This mutagenesis was performed with primers XhoSalXaHTPGR 5'gcttatcgataccctcgaggaaggcgatatcg3' and XhoSalXaHTPGF 5'cgatatcgccttcctcgagggtatcgataagc3', yielding pBKS/XALB3XhoI (step 6). Finally, the *Xho*I cassette of pBC/Op3-4/XALB2 was subcloned into the *Sal*I restriction site of pBKS/XALB3XhoI, yielding pBKS/Op3-4/XALB2-3 (step 7). This construct harbours an *Xho*I cassette containing complete operons 3 and 4 from XALB1, *albXXI* from XALB2 and *albXXII*

from XALB3. An *Xho*I site was added to the *Bam*HI site of the pLAFR3 shuttle vector polylinker using the adaptor AdApTBamHIXhoI 5'gatcgctcgagc3', yielding pLAFR3XhoI (step 8). The *Xho*I cassette from pBKS/Op3-4/XALB2-3 was then cloned into pLAFR3XhoI, yielding pOp3-4/XALB2-3 (step 9). This last construct was used, along with pALB571 (operons 1 and 2), for heterologous expression of albicidin in *X. axonopodis* pv. vesicatoria.

Albicidin production assays

[00156] The four combinations of plasmids (i.e. pUFR043–pLAFR3, pUFR043-pOp3-4/XALB2-3, pAlb571-pLAFR3 and pAlb571-pOp3-4/XALB2-3) were transferred into *X. axonopodis* pv. vesicatoria strain Xcv 91-11BR1 by triparental mating. Exconjugant clones resistant to tetracycline and kanamycin were isolated. Assays for albicidin production were performed with these exconjugants clones using the same method described in Example 1 except that tetracycline (12 mg/ml) and/or kanamycin (50 mg/ml) were added to SPA medium. Tetracycline and kanamycin resistant *E. coli* clones, DH5αKT and DH5αAlb^{r}KT (Table 1), were used as tester strains to evaluate albicidin production to ensure that growth inhibition was not due to the presence of these two antibiotics in SPA medium. Both clones, DH5αKT and DH5αAlb^{r}KT, are tetracycline and kanamycin resistant because they carry plasmids pLAFR3 and pUFR043. The albicidin resistant DH5αAlb^{r}KT clone derived from strain DH5αAlb^r (Table 1) which is a spontaneous albicidin resistant clone isolated in a growth inhibition zone produced by *X. albilineans* strain Xa23R1.

[00157] Without antibiotics in the SPA medium, growth of clones DH5αKT and DH5αAlb^{r}KT was not inhibited in all assays performed with the different *X. axonopodis* pv. vesicatoria exconjugants. Surprisingly, when kanamycin was present in the SPA medium, growth of both DH5αKT and DH5αAlb^{r}KT was inhibited in all assays performed with the *X. axonopodis* pv. vesicatoria exconjugants. These results suggested that, in the presence of kanamycin, all *X. axonopodis* pv. vesicatoria exconjugants produced an antibiotic inhibiting growth of *E. coli*. Because exconjugants containing only empty vectors (pUF043 and pLAFR3) induced inhibition of *E. coli*, this antibiotic did not result from the expression of XALB1, XALB2 and/or XALB3. Additionally, there was no cross resistance between this antibiotic and albicidin. When tetracyclin was present in the bioassay medium, but not kanamycin, growth of the albicidin resistant clone (DH5αAlb^{r}KT) was not inhibited by any of the exconjugants. In contrast, growth of the albicidin susceptible *E. coli* strain (DH5αKT) was inhibited by the exconjugants harbouring pALB571 and pOp3-4/XALB2-3 plasmids, but not by exconjugants harbouring the other three combinations of plasmids (Table 10). This result suggested that expression of the XALB1, XALB2 and XALB3 regions in *X. axonopodis* pv. vesicatoria (harbouring pALB571 and pOp3-4/XALB2-3 plasmids) led to the production of an albicidin-like antibiotic. This product inhibited growth of an albicidin sensitive *E. coli* (DH5αKT) and had no effect on the growth of an albicidin resistant clone

(DH5αAlb^{r}KT).

[00158] Preliminary results indicated that pLAFR3 derived plasmids were relatively unstable in the absence of tetracycline in the culture medium, suggesting that genes carried by pOp3-4/XALB2-3 were not expressed when *X. axonopodis* pv. vesicatoria exconjugants pALB571/pOp3-4/XALB2-3 were grown without tetracycline. Consequently, these exconjugants did not produce the albicidin-like compound in absence of any antibiotic in the culture medium (Table 10). Preliminary results also indicated that pUFR043 derived plasmids are relatively stable in *X. axonopodis* pv. vesicatoria in absence of antibiotic selection, suggesting that genes carried by pALB571 are expressed when *X. axonopodis* pv. vesicatoria exconjugants pALB571/pOp3-4/XALB2-3 were grown on media without kanamycin. Consequently, these exconjugants produced the albicidin-like compound on SPA containing only tetracyclin.

[00159] Two *E. coli* DH5αKT clones, that spontaneously grew within the growth inhibition zone of a *X. axonopodis* pv. vesicatoria pALB571-pOp3-4/XALB2-3 exconjugant on SPA + tetracycline medium, were isolated and tested for resistance to albicidin. No growth inhibition was observed when these clones were used as tester strains in an albicidin production assay performed with *X. albilineans* Xa23R1. These results showed that cross-resistance occurs between the albicidin-like product of *X. axonopodis* pv. vesicatoria and albicidin produced by *X. albilineans*, suggesting that both molecules are similar. Comparison of chemical characteristics of the two molecules will, however, be necessary to confirm that the two molecules are identical.

[00160] The invention includes the isolation and sequencing of a region of 55,839 bp from *X. albilineans* strain Xa23R1 containing the major gene cluster XALB1 involved in albicidin production. Analysis of this region allowed us to predict the genetic organization of the gene cluster XALB1 which contains 20 ORFs grouped in four or five operons (Figure 1). Because *albXVIII* is a truncated gene, XALB1 genes may be organized in five operons. Therefore, we will from now on consider *albXVII* as part of operon 4 and *albXIX* and *albXX* as part of operon 5. Similar operon-type organizations for antibiotic biosynthesis clusters are well known and have been postulated to facilitate cotranslation of genes within the operon to yield equimolar amounts of proteins for optimal interactions to form the biosynthesis complexes (Cane, 1997). Overlapping genes involved in the same process are also quite common in bacteria (Normark *et al.*, 1983).

[00161] Previous results of transposon mutagenesis and complementation studies (Rott *et al.*, 1996; Rott, unpublished results) are in accordance with the predicted genetic organization of XALB1 described in this study, and allowed us to establish that operons 1, 2 and 3 are involved in albicidin biosynthesis: (I) Tox$^-$ mutants with a Tn*5-gusA* insertion site located in DNA fragments B, C, G and D were complemented by cosmid pALB571 and not by cosmid pALB540, confirming that cosmid pALB571 potentially contains the entire operon 1; (ii) Tox$^-$ mutants with a Tn*5-gusA* insertion site located in DNA fragments A and H were complemented

by both cosmids pALB540 and pALB571, confirming that both cosmids potentially contain the entire operon 2; (iii) mutant XaAM1 with a Tn*5*-*gusA* insertion site located in DNA fragment J is the only Tn*5* Tox⁻ mutant complemented by cosmid pALB540 and not by cosmid pALB571, confirming that cosmid pALB540 potentially contains the entire operon 3. Our mutagenesis studies did not confirm that operons 4 and 5 are required for biosynthesis of albicidin. The para-aminobenzoate (PABA) is required for the growth of many bacteria probably including *X. albilineans*, suggesting that a mutation in *albXVII* may be lethal and explaining why we did not obtain any mutant disrupted in this gene.

[00162] Putative bidirectional promoters were identified between operons 1 and 2 (Huang et al., 2001) and between 3 and 4 (Figure 7), confirming the prediction of genetic organization of XALB1. The region upstream from operon 1 is 100 % identical to the region upstream from the *xabB* start codon which was described as a functional promoter during the phase of albicidin accumulation in Australian strain Xa13 of *X. albilineans* (Huang *et al.*, 2001). Involvement of several operons in albicidin biosynthesis suppose that they are transcribed during the same time. Promoter activities of regions upstream from putative operons 2, 3, 4 and 5 need to be determined to precise if these promoters are functional during the same growth phase of *X. albilineans* as the promoter upstream from operon 1.

[00163] Potential rho-independent transcription terminators were identified downstream from operons 1, 3 and 4 (Figure 6) confirming prediction of the genetic organization of these three operons. Because operons 2 and 5 are convergent (Figure 1) and separated by a very short region of 22 bp between *albIX* and *albXX*, stop codons may allow transcription termination in the absence of sequences corresponding to potential rho-independent transcription terminators downstream from these operons. It is quite possible that simultaneous transcription of operons 2 and 5 involving the presence of two RNA polymerases (one on each strand of DNA) may cause RNA polymerases to pause at the end of each operon because of steric interference between both polymerase complexes in the same short region.

[00164] The presence of putative RBSs upstream of the ATG start codons of all ORFs, except for *albXVIII*, suggests that these ORFs are translated in *X. albilineans*. The absence of a canonical RBS upstream from *albXVIII* further indicates that this ORF is probably not expressed. GTG and TTG codons (usually valine and leucine codons) generally serve as procaryotic start codons when located near the 5' end of an RNA message, but GTG start codons were also described far from the 5' end of messenger RNA in the bacitracin biosynthesis cluster of *B. lucheniformis* (Genbank Accession No. AF184956) or in the bleomycin biosynthetic gene cluster of *S. verticillus* (Genbank Accession No. AF210249). This is in accordance with the fact that the two potential TTG start codons are the first start codons in operons 1 and 4 of XALB1, and that the two potential GTG start codons initiate internal cistrons. The *albI* and *albXVII* genes,

like *xabB* (Huang *et al.*, 2001), use TTG as a start codon, which may impose post-transcriptional control of the rate of gene product formation (McCarthy and Gualerzi, 1990).

[00165] The predicted genetic organization of operons 1 and 2 presents similarities with the organization of the region involved in albicidin production in strain Xa13 of *X. albilineans* from Australia (Huang *et al.* 2000a, Huang *et al.*, 2001). This latter region also contains two divergent operons involved in albicidin production, one comprising the *xabB* gene (similar to *albI*, but with a large deletion) and the *xabC* gene (100% identical to *albII*) and the other containing *thp* gene (100% identical to *albV*). In addition, the sequence between the two operons in strain Xa13 is 100% identical to the sequence between operons 1 and 2, indicating that both clusters are controlled by the same bidirectional promoter. However, transposon mutagenesis studies of Xa13 showed no evidence of another cistron downstream of *xabC* that may be involved in albicidin production (Huang *et al.*, 2000a), suggesting that the Xa13 *xab* operon differs from the Xa23R1 operon 1, which contains two additional genes downstream from *albII* that are potentially involved in albicidin production (*albIII* and *albIV*; refer Figure 1).

[00166] Homology analysis revealed that four NRPS and/or PKS genes are present in XALB1 (Figure 2), and these genes may be involved in the biosynthesis of the albicidin polyketide-polypeptide backbone (*albI*, *albIV*, *albVII* and *albIX*). NRPS and PKS enzymes are generally organized into repeated functional units known as modules, each of which is responsible for a discrete stage of polyketide or polypeptide chain elongation (Cane and Walsh, 1999). Each PKS or NRPS module is made up of a set of three core domains, two of which are catalytic and one of which acts as a carrier, and together are responsible for the central chain-building reactions of polyketide or polypeptide biosynthesis. Both PKS and NRPS core domains utilize analogous acyl-chain elongation strategies in which the growing chain, tethered as an acyl-S-enzyme to the flexible 20 Å long phosphopantetheinyl arm of an acyl carrier protein (ACP) or peptidyl carrier protein (PCP) domain, acts as the electrophilic partner that undergoes attack by a nucleophilic chain-elongation unit, a malonyl- or aminoacyl-S-enzyme derivative, respectively, itself covalently bound to a downstream ACP/PCP domain. In the case of a PKS, the fundamental chain-elongation reaction, a C-C bond-forming step, is mediated by a ketosynthase (KS) domain that catalyzes the transfer of the polyketide acyl chain to an active-site cysteine of the KS domain, followed by condensation with the methylmalonyl- or malonyl-S-ACP by a decarboxylative acylation of the malonyl donor unit. An additional essential component of the core PKS chain-elongation apparatus is an associated acetyltransferase (AT) domain, which catalyzes the priming of the donor ACP sidearm with the appropriate monomer substrate, usually methylmalonyl- or malonyl-CoA. The comparable core domains of an NRPS biosynthetic module function in a chemically distinct but architecturally and mechanistically analogous fashion. In the latter case, the key chain-building reaction, a C-N bond-forming reaction, involves the generation of the characteristic peptide bond by nucleophilic attack of the amino group of an amino acyl-S-PCP

donor on the acyl group of an upstream electrophilic acyl- or peptidyl acyl-S-PCP chain, catalyzed by a condensation (C) domain. In functional analogy to the PKS AT domain, the core of the NRPS module utilizes an adenylation (A) domain to activate the donor amino-acid monomer as an acyl-AMP intermediate, which is then loaded onto the downstream PCP side chain. Both the AT and A domains of the respective PKS and NRPS modules act as important gatekeepers for polyketide or polypeptide biosynthesis, exhibiting strict or at least high specificity for their cognate malonyl-CoA, methylmalonyl-CoA or amino acid substrates. In addition to the basic subset of core domains, each PKS or NRPS also has a special set of dedicated domains responsible both for the initiation of acyl-chain assembly by loading of a starter unit onto the first, furthest upstream PKS/NRPS module, as well as a chain-terminating thioesterase (TE) domain, most often found fused to the last module, that is responsible for detachment of the most downstream covalent acyl enzyme intermediate and off-loading of the mature polyketide or polypeptide chain (Cane and Walsh, 1999).

[00167] XALB1 potentially encodes four PKS modules and seven NRPS modules. Most of the bacterial NRPS gene clusters described up to now are organized in operon-type structures, encoding multi modular NRPS proteins with individual modules organized along the chromosome in a linear order that parallels the order of amino acids in the resultant peptide, following the "colinearity rule" for the NRPS-template assembly of peptides from amino acids (Cane, 1997; Cane *et al.*, 1998; Cane and Walsh, 1999; von Döhren *et al.*, 1999). PKS and NRPS modules are apparently not organized according to this "colinearity rule" for albicidin biosynthesis because of the following features : (I) NRPS and PKS genes are expressed in two divergent operons; (ii) no AT domain was identified in PKS-2 and PKS-3 domains, suggesting involvement of a separate enzyme ; (iii) the A domain of NRPS-2 is not functional, suggesting the involvement of a *trans*-acting A domain ; (iv) a single chain-terminating TE domain was identified in XALB1 which may be responsible of the release of the full length albicidin polyketide-polypeptide backbone from the enzyme complexes. Exception to the "colinearity rule" has also been shown for the syringomycin synthetase of *P. syringae* (Guenzi *et al.*, 1998), for the exochelin synthetase of *Mycobacterium smegmatis* (Yu *et al.*, 1998) and for the bleomycin synthetases of *S. verticillus* (Du *et al.*, 2000).

[00168] On the basis of the deduced functions of individual NRPS and PKS domains we have aligned the four PKS and the seven NRPS modules to suggest two different putative linear models for the synthesis of the albicidin polyketide-peptide backbone (Figure 9). In the Figure, NRPS and PKS domains are abbreviated as follows: A, adenylation; ACP, acyl carrier protein; AL, acyl-CoA ligase; AT, acyltransferase; C, condensation; HBCL,hydroxybenzoate-CoA ligase; KR, ketoreductase; KS, ketoacyl synthase; PCP, peptidyl carrier protein. Asn designates asparagine. X1 and X2 indicate substrates incorporated by NRPS - 1 and 3 and by NRPS-6 and 7, respectively. The crossed A domain in NRPS-2 indicates that this

deleted domain may be not functional. In model 1, (Figure 9A), (i) the PKS-1 module alone is responsible for the initiation of the acyl-chain assembly, (ii) PKS-4 (HBCL) interacts with PKS-2 and PKS-3 as an AT domain to allow acyl transfer and (iii) NRPS-5 interacts with only NRPS-2. In model 2 (Figure 9B) two different modules, PKS-1 and PKS-4, are responsible for this initiation step. Model 2 leads to the biosynthesis of four different polyketide-polypetide backbones; in this model (i) PKS-1 (AL) and PKS-4 (HBCL) are in competition for initiation of albicidin precursors; (ii) a separate AT enzyme (potentially AlbXIII) interacts with PKS-2 and PKS-3 to allow acyl transfer; (iii) NRPS-5 interacts with NRPS-2; and (iv) NRPS-5 and NRPS-6 are in competition for interaction with NRPS-4.

[00169] Both models are based on the fact that PKS-1 contains the AL and ACP1 domains, and PKS-4 shows homology with the hydroxybenzoate-CoA ligases. In other PKS systems, an N-terminal AL domain is involved in the activation and incorporation of an 3,4-dihydroxycyclo hexane carboxylic acid, a 3-amino-5-hydroxybenzoic acid or a long-chain fatty acid as a starter (Aparicio *et al.*, 1996; Motamedi and Shafiee, 1998; Tang *et al.*, 1998; Duitman *et al.*, 1999). PKS-4 may be also involved in the activation and incorporation of hydroxy-benzoate but this latter domain lacks any ACP or PCP domain, suggesting that PKS-4 is responsible for initiation of the acyl-chain assembly (Figure 9B) onto one of the three ACP domains of AlbI (ACP1, ACP2 or ACP3). The 277 amino-acids preceding the PKS-4 module in AlbVII may be necessary for the intercommunication between AlbVII and AlbI. The presence of two different PKS modules potentially involved in the initiation of the acyl-chain assembly suggests a competition of these two modules for the initiation of two different albicidin polyketide-polypeptide backbones, and this could contribute to the production of multiple, structurally related albicidins by the same cluster XALB1. Production of two different components, one initiated by PKS-4 containing an additional aromatic ring due to incorporation of hydroxybenzoate, may explain why partial characterization of albicidin indicated the presence of a variable number (three or four) of aromatic rings (Huang *et al.*, 2001).

[00170] In AlbI, PKS-1 is followed by the PKS-2 module which contains a KS domain and a KR domain upstream from two ACP domains (ACP2 and ACP3) and it lacks any discernable AT domain. Tandem ACP domains are unusual within PKS modules but have been shown to occur in the biosynthesis of several fungal and bacterial polyketide synthases (Mayorga and Timberlake, 1992; Yu and Leonard, 1995; Takano *et al.*, 1995; Albertini *et al.*, 1995). However, the significance of the tandem ACP domains in these systems has not been solved yet. In our model 2, one of the tandem ACP (ACP2 or ACP3) may interact with PKS-4 for the initiation of an acyl-chain assembly (Figure 9B). The absence of an AT domain in the PKS-2 module suggests that a separate AT domain is indispensable for the elongation of the acyl-chain initiated by this module. Separate AT enzymes encoded elsewhere in the genome were described in other systems for two PKS modules lacking AT domains: malonyl-CoA transacyclase gene

(*fen*F) located immediately upstream from the *B. subtilis* PKS-NRPS *mycA* gene (Duitman *et al.*, 1999) and an AT gene located 20kb upstream from the *M. xanthus* NRPS-PKS *tal* gene (Paitan *et al.*, 1999). We have not identified an AT gene in the gene cluster XALB1 and in the two other genomic regions involved in albicidin production, XALB2 and XALB3, suggesting that the *trans*-acting AT gene may be encoded elsewhere in the genome. However, AlbXIII, which contains the motif GHSxG conserved in AT domains, may be potentially involved in the acyl transfer, but the similarity of AlbXIII with AT domains is not high enough to confirm this potential function of AlbXIII (Figure 10). Figure 10A describes alignment of the conserved motifs in AT domains from RifA-1, -2, -3, RifB-1, RifE-1 (Rifamycin PKSs, August *et al.*, 1998) and BlmVIII (Bleomycin PKS; Du *et al.*, 2000), identical amino acids are shown in bold. Figure 10B describes alignment of AlbXIII (SEQ ID NO. 38), FenF (a malonyl-CoA transacylase located upstream from *mycA*, Duitman *et al.*, 1999) and LipA (a lipase; Valdez *et al.*, 1999); amino acids identical to conserved AT domains motifs are shown in bold.

[00171] AlbXIII contains only four of the eleven amino acids conserved in AT domains of rifamycin PKSs (August *et al.*, 1998) and Bleomycin PKS (Du *et al.*, 2000), and the AlbXIII sequence appears to be more closely related to lipases such as LipA (Valdez *et al.*, 1999) rather than to AT domains (Figure 10). However, FenF, the *trans*-acting AT domain involved in mycosubtilin biosynthesis, contains only seven of the eleven amino acids conserved in AT domains (Duitman *et al.*, 1999; Figure 10). AlbVII, that contains a HBCL domain, may be another candidate for the acyl transfer in PKS-2 (Figure 9A) because HBCL exhibits some similarity with A domains at the level of cores A1, A2, A3, A4, A5 and A6 (Table 6). However, no HBCL involved in such a function has been described in the PKSs characterized so far.

[00172] In AlbI, PKS-2 is followed by the PKS-3 module which contains the KS2 and the PCP1 domains and it lacks any discernable AT or A domain. PKS-3 is located upstream from the NRPS modules and should therefore be involved in the linkage of polyketide and polypeptide moieties. The presence of a PCP domain in the PKS-3 module suggests the involvement of a *trans*-acting A domain rather than an AT domain. A putative candidate for this *trans*-acting A domain is the AlbIV NRPS-5 A domain because of the lack of a C domain in the AlbIV NRPS-5 module. However, by analogy with the BlmVIII PKS module, which is involved in the linkage of polypeptide and polyketide moieties of bleomycin and which contains an AT domain followed by a PCP domain (Du *et al*, 2000), the presence of a PCP is not incompatible with a possible interaction of the AlbI PKS-3 module with a separate AT domain. This latter *trans*-acting AT domain may be the same that interacts with the AlbI PKS-2 module, the AlbVII PKS-4 module, AlbXIII or an unindentified separate AT domain.

[00173] In AlbI, the PKS-3 module is followed by four NRPS modules. The NRPS-1, 2 and 3 modules display the ordered C, A and PCP domains, suggesting that they are

involved in the incorporation of three amino acid residues. The A domain of the NRPS-2 module exhibits poor consensus at A2, A3, A5, A7, A8 A9 and A10 motifs and lacks completely the A6 motif (Table 6). In addition the NRPS-2 substrate binding pocket is partially deleted (Figure 5). These features strongly suggest that the NRPS-2 A domain is inactive and that the loading of an amino-acid on the NRPS-2 PCP domain (PCP3) is possibly catalyzed by a *trans*-acting A domain as in HMWP1 (Gehring *et al.*, 1998) and BlmIII (Du *et al.*, 2000). A putative candidate for this *trans*-acting A domain is the NRPS-5 A domain present in AlbIV because of the lack of a C domain in NRPS-5 (Figure 2). The additional sequence of 300 amino-acids present in the A domain of NRPS-5 may be necessary for the intercommunication between AlbI and AlbIV. As a consequence of the interaction between NRPS-2 and NRPS-5, a competition between PCP-3 and PCP-5 domains must occur to bind the amino acid activated by the NRPS-5 A domain. A similar competition between two PCP domains was described for syringomycin biosynthesis, during the interaction between SyrB, which contains A and PCP domains, and the last module of SyrE which contains C and PCP domains (Guenzi *et al.*, 1998). The NRPS-4 module contains only a C domain which may transfer the intermediate products synthetized by AlbI to a PCP domain present in an other albicidin synthase. Similar transfers were described for mycosubtilin biosynthesis in which the MycA and MycB C-terminal C domains interact with the MycB and MycC N-terminal A domains, respectively (Duitman *et al.*, 1999). Two different PCP domains may be involved in the transfer of the intermediate products synthetized by AlbI: the PCP-5 and PCP-6 domains which are present in the AlbIV NRPS-5 and AlbIX NRPS-6 modules, respectively. This possible competition between the two NRPS modules that contain two different A domains could also contribute to the production of multiple, structurally related albicidins by the gene cluster XALB1 (Figure 9B). Because of the absence of a C-domain in the AlbIX NRPS-6 module, the intermediate product bound on the AlbIV PCP-5 domain would be necessarily transferred to the AlbIX PCP-7 domain, like the intermediate product bound on AlbIX PCP-6. AlbIX NRPS-7, which contains the single chain-terminating TE domain, may then be responsible for detachment of the mature albicidin polyketide-polypeptide backbone from the complex of enzymes.

[00174] The linear model 1 implies that operon 1 and operon 2 in *X. Albilineans* strain Xa23R1 from Florida potentially produce only one albicidin polyketide-polypetide backbone, with a competion at the level of ACP2/ACP3 and PCP3 and PCP5 which could explain the production by *X. albilineans* of compounds structurally related to albicidin (Figure 9A). The linear model 2 implies that operon 1 and operon 2 in *X. albilineans* strain Xa23R1 from Florida potentially produce four different albicidin polyketide-polypetide backbones (Figure 9B) because of (i) the competition of AL and HBCL domains for initation of acyl chain assembly and (ii) the competition of AlbIV NRPS-5 and AlbIX NRPS-6 modules for the incorporation of the next to last amino acid of the albicidin backbone. These four albicidin backbones may lead to the

production of four components structurally very different. The polyketide moieties of the acyl chains initiated by the AlbI AL domain or by the AlbVII HBCL domain may be very different. The polyketide moiety of acyl chains initiated by the AlbVII HBCL domain may be shorter and may contain an additional aromatic ring. The presence of four structuraly different metabolites may explain the difficulty observed by Birch and Patil (1985a) to purify albicidin and to determine its chemical structure.

[00175] Homology analysis also revealed that AlbI NRPS-1 and 3 and AlbIX NRPS-6 and 7 specify unusual substrates which seem to contain an amino group and a carboxylate group but to be different from α-amino acids and β-alanine. Identification of several aromatic rings in albicidin (Huang *et al.*, 2001) suggested that NRPS-1, -3, -6 and -7 are involved in incorporation of aromatic substrates. By analogy with the Asp235Val alteration in the β-Ala specificity-conferring code (Du *et al.* 2000), the Asp235Ala alteration in the NRPS-1, -3, -6 and -7 signatures could be consistent with a large distance between the amino group and the carboxylate group in the substrate specified by these modules. Based on this hypothesis, we suggest that operons 3, 4 and 5 are involved in the biosynthesis of two aromatic substrates: the para-aminobenzoate potentially synthesized by AlbXVII (para-aminobenzoate synthase), and the carbamoyl benzoate potentially synthesized by AlbXX (hydroxybenzoate synthase) and AlbXV (carbamoyl transferase). Incorporation of these nonproteinogenic substrates may explain why albicidin is insensitive to proteases (Birch and Patil, 1985a).

[00176] According to biosynthesis model 1 leading to the biosynthesis of only one polyketide-polypeptide albicidin backbone that may correspond to the major component produced by XAlb1, we propose a model allowing prediction of the composition and the structure of albicidin (Figure 11). In the Figure, NRPS and PKS domains are abbreviated as follows: A, adenylation; ACP, acyl carrier protein; AL, acyl-CoA ligase; C, condensation; KR, ß-ketoreductase; KS, ß-ketoacyl synthase; PCP, peptidyl carrier protein. C atoms of albicidin-backbone are numbered 1 to 38. Bold methyl groups correspond to methylation of the albicidin backbone by AlbII or AlbVI. In this model, albicidin biosynthesis is initiated by loading of an acetyl-CoA by PKS-1 (step 1), and the chain product is elongated by incorporation of (I) malonyl-CoA by PKS-2 and PKS-3 (steps 2 and 3), (ii) para-aminobenzoate or carbamoyl benzoate by NRPS-1 and NRPS-3 (steps 4 and 6), (iii) asparagine by NRPS-2 coupled to NRPS-5 (step 5) and (iv) para-aminobenzoate or carbamoyl benzoate by NRPS-6 and NRPS-7 (steps 7 and 8). The presence of the KR domain in the PKS-2 module may lead to the formation of an hydroxyl group at the C_2 atom of the albicidin backbone. This hydroxyl group might be methylated by AlbVI (*O*-methyltransferase). The acyl chain may also be modified by AlbII (*C*-methyltransferase) at C_{13} or C_{14}.

[00177] The chemical composition ($C_{40}O_{15}N_6H_{35}$), the molecular weight (839), and the structure of the putative XALB1 product are in accordance with the partial characterization of albicidin published by Birch and Patil (1985a) which indicated that albicidin contains approximately 38 carbon atoms and a carboxylate group and that the molecular weight of albicidin was about 842. The presence of two ester linkages in our predicted albicidin structure is also in accordance with the fact that albicidin is detoxified by the AlbD esterase (Zhang and Birch, 1997). However, an unpublished albicidin analysis cited by Huang *et al.* (2001) indicated the presence of (I) two OCH3 groups and not one as in our predictive albicidin structure, (ii) one CN linkage and not eleven as in our predictive albicidin structure and (iii) a trisubstituted double bond that is not present in the putative XALB1 product.

[00178] In conclusion, homology analysis of XALB1 revealed unprecedented features for hybrid polyketide-peptide biosynthesis in bacteria involving a *trans*-action of four PKS and seven NRPS separate modules which could contribute to the production of multiple, structurally related polyketide-peptide compounds by the same gene cluster. Characterization of the full chemical structure of albicidin may be necessary to validate these models. Four NRPS modules seem to activate a very unusual substrate. Over- expression and purification of A domains from these four NRPS modules will be necessary to examine their substrate specificities. Substrate specificity of each A domain will therefore be determinated by analysis of the ATP-PPi exchange reaction with different substrate putatively incorporated into albicidin. Investigating albicidin backbone biosynthesis will be of great interest because such information adds to the limited knowledge as to how PKS and NRPS interact and how they might be manipulated to engineer novel molecules, and may explain how *X. albilineans* produces several structurally related, toxic compounds.

[00179] Cloning and sequencing of XALB2 showed that the same phosphopantetheinyl transferase is required for albicidin production in an *X. albilineans* strain from Florida and in an *X. albilineans* strain from Australia (Huang *et al.*, 2000b), explaining the precedented results showing that strain LS156 mutated in *xabA* (100% identical to *albXXI*) was not complemented by pALB540, pALB571 and pALB639 (Rott *et al.*, 1996). Mutant LS156 was shown to be complemented by a construction containing the coding sequence of *xabA* in fusion with *lacZ*, revealing that *xabA* is required for albicidin production and that no other cistron downstream from *xabA* was involved in albicidin production (Huang *et al.*, 2000b). However, this complementation study did not allow determination of whether *xabA* is transcribed as a part of a larger operon. Here we disclose the complementation of mutant AM37 with a 2986 bp insert from *X. albilineans* containing *albXXI* (100% identical to *xabA*), confirming that *albXXI* is involved in albicidin biosynthesis and indicating that the promoter of *albXXI* is present in the 2986 bp insert and that *albXXI* is not expressed as part of a operon.

[00180] Cloning and sequencing of XALB3 showed that a heat shock protein HtpG was involved in albicidin production in *X. albilineans*. The heat shock protein HtpG is an *Escherichia coli* homologue of eukaryotic HSP90 molecular chaperone. Hsp90 from eukaryotes has been demonstrated to possess chaperone activity (Jakob *et al.*, 1995), acting as a non-ATP dependent "holder," and it also has an important role in signal transduction and the cell cycle. This protein is essential in both drosophila and yeast (Borkovich et al., 1989; Cutforth and Rubin, 1994). In contrast, the HtpG gene can be deleted in *E. coli* with no effect on the viability of the strain with the exception of decreased growth rate at high temperatures (Bardwell and Craig, 1988). The *in vivo* role of the HtpG protein remains unknown. However, preliminary results indicated that HtpG facilitates *de novo* protein folding in stressed *E. coli* cells, presumably by expanding the ability of the DnaK-DnaJ-GrpE molecular chaperone system to interact with newly synthesized polypeptides (Thomas and Baneyx, 2000). Furthermore, HtpG was copurified in *E. coli* with MccB17 synthetase, an enzyme involved in the biosynthesis of the peptide antibiotic microcin B17 which inhibits DNA replication by induction of the SOS repair system, suggesting the requirement of HtpG for production of the antibiotic (Li *et al.*, 1996). However, when microcin B17 production by the *E. coli* strain deleted for HtpG was compared to the one of the parental strain, there was no effect on microcin B17 production *in vivo*. This result implied that the copurification of HtpG with the MccB17 synthetase was potentially an artifact, or that another *E. coli* chaperone could substitute for HtpG (Milne *et al.*, 1999). To examine the effect of HtpG on the reconstitution of MccB17 synthetase *in vitro*, the chaperone was expressed and purified as a fusion to a hexahistidine (His_6) tag. Addition of the His_6-HtpG did not stimulate MccB17 synthetase reconstitution or heterocyclisation activity *in vitro*, suggesting that HtpG mediates complex assembly or stabilizes protein subunits prior to the hetero-oligomerisation (Milne *et al.*, 1999). Based on these results, we suggest that the function of AlbXXII is to mediate complex assembly by facilitating *de novo* protein folding of PKS and NRPS enzymes (AlbI, AlbIV, AlbVII and AlbIX) involved in the albicidin backbone biosynthesis.

[00181] Characterization of the complete sequence of XALB1, XALB2 and XALB3 clusters enables one to characterize all enzymes of the albicidin biosynthesis pathway including structural, resistance, secretory and regulatory elements, and to engineer overproduction of albicidin. For example one may insert expression enhancing DNA into the genome of *X. albilineans* in a position operable to enhance expression of the Albicidins Biosynthesis Gene Clusters. One may also modify naturally occurring Albicidins to obtain additional non-naturally occurring antibiotics by adding DNA encoding additional enzymes selected to produce a modified albicidin like molecule. This approach will allow (I) the purification of albicidin and the other compounds structurally related and potentially produced by the same biosynthesis apparatus; (ii) the characterization of chemical structure of albicidin; (iii) the investigation of mode of action of

albicidin in the pathogenesis of *X. albilineans* in sugarcane; and (iv) the characterization of the bactericidal activity of albicidin. For example one may also increase the resistance of plants to damage from *X. albilineans* infection by inserting one or more of the resistance genes identified herein into the genome of the plant. One may also provide materials to prevent damage by albicidin produced by *X. albilineans* by applying an agent that blocks expression of the Albicidin Biosynthesis Gene Clusters to the plant to be protected. One may also use portions of the DNA of the Albicidin Biosynthesis Gene Clusters to obtain agents useful in blocking expression of albicidin by screening materials against a modified hast cell line that expresses the Albicidin Biosynthesis Gene Clusters and selecting for materials that stop or decrease albicidin production.

Table 1 : Bacterial strains and plasmids used in this study

	Relevant characteristics[a]	Reference or source
Strains		
E. coli		
DH5α	F-/80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *deo*R *recA*1 *endA*1 *hsdR17*(r_k^- m_k^+) supE44 thi-1 *gyrA*96 *relA*1	Gibco-BRL
DH5α MCR	DH5α*mcrA mcrBC mrr*	"
Xcv 91-11B	Wild type strain of *Xanthomonas axonopodis* pv. vesicatoria from tomato (race 3)	Astua-Monge *et al.*, 2000
Xcv 91-11BR1	Spontaneous Rif[r] derivative of Xcv 91-11B	This study
DH5αKT	*Escherichia coli* DH5α strain transformed by both pUFR043 and pLAFR3 plasmids	"
DH5αAlb[r]	Spontaneous Alb[r] derivative of DH5α	"
DH5αAlb[r]KT	DH5αAlb[r] transformed by both pUFR043 and pLAFR3 plasmids	"
Plasmids		
PBR325	Tc[r], Ap[r], Cm[r]	Gibco-BRL
pBCKS (+)	Cm[r]	Stratagene
pBluescript II KS (+)	Ap[r]	"
PRK2073	PRK2013 derivative, Km[s] (*npt*::Tn*7*), Sp[r], Tra[+], helper plasmid	Leong *et al.*, 1982
pUFR043	IncW Mob[+] LacZα Gm[r], Km[r], Cos	De Feyter and Gabriel, 1991
pAlb540	47 kb insert from Xa23R1 in pUFR043, Gm[r], Km[r]	Rott *et al.*, 1996
pAlb571	36.8 kb insert from Xa23R1 in pUFR043, Gm[r], Km[r]	"
PAlb639	36 kb insert from Xa23R1 in pUFR043, Gm[r], Km[r]	"
pAM15.1	24 kb *Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM15 in pBR325, Km[r], Tc[r], Ap[r], Cm[r]	"
pAM40.2	11 kb *Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM40 in pBR325, Km[r], Tc[r], Ap[r], Cm[r]	"
pAM45.1	12 kb *Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM45 in pBR325, Km[r], Tc[r], Ap[r], Cm[r]	"
pAM12.1	13 kb *Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM12 in pBR325, Km[r], Tc[r], Ap[r], Cm[r]	A
PAM36.2	9 kb*Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM36 in pBR325, Km[r], Tc[r], Ap[r], Cm[r]	A
pAlb389	37 kb insert from Xa23R1 in pUFR043, Gm[r], Km[r]	This study
pAC389.1	2.9 kb insert from Xa23R1 in pUFR043, Gm[r], Km[r]	"
PAlb639A	9.4 kb insert from Xa23R1 in pUFR043, Gm[r], Km[r]	"
PEV639	2.6 kb *Sal* I insert from Xa23R1 in pUFR043, Gm[r], Km[r]	"
pBC/A'	7.5 kb *Kpn* I fragment carrying a part of fragment A from pAlb571 in pBCKS (+), Cm[r]	"
pBC/AF	15.2 kb *Eco*R I fragment carrying fragments A and F from pALB540 in pBCKS (+), Cm[r]	"
pBC/B	11.0 kb *Kpn* I fragment B from pAlb571 in pBCKS (+), Cm[r]	"
pBC/C	6.0 kb *Kpn* I fragment C from pAlb571 in pBCKS (+), Cm[r]	"
pBC/E	2.8 kb *Kpn* I fragment E from pAlb571 in pBCKS (+), Cm[r]	"
pBC/F	2.5 kb *Kpn* I-*Eco*R I fragment F from pAlb571 in pBCKS (+), Cm[r]	"
pBC/G	1.9 kb *Eco*R I fragment G from pAlb571 in pBCKS (+), Cm[r]	"

Table 1 : Bacterial strains and plasmids used in this study

	Relevant characteristics[a]	Reference or source
pBC/I	1.4 kb *Kpn* I-*Eco*R I fragment I from pAlb571 in pBCKS (+), Cm^r	"
pBC/J	0.6 kb *Eco*R I fragment J from pALB540 in pBCKS (+), Cm^r	"
pBC/K	4.7 kb *Eco*R I fragment K from pALB540 in pBCKS (+), Cm^r	"
pBC/L	0.4 kb *Eco*R I fragment L from pALB540 in pBCKS (+), Cm^r	"
pBC/N	7.7 kb *Eco*R I fragment N from pALB540 in pBCKS (+), Cm^r	"
pUFR043/D=	2.2 kb *Eco*R IB*Sau*3A I fragment carrying a part of fragment D from pAlb571 in pUFR043	"
pAM1	5 kb *Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM1 in pBluescript II KS (+), Km^r, Ap^r	"
pAM4	12 kb *Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM4 in pBluescript II KS (+), Km^r, Ap^r	"
pAM7	6 kb *Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM7 in pBluescript II KS (+), Km^r, Ap^r	"
pAM10	7 kb *Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM10 in pBluescript II KS (+), Km^r, Ap^r	"
pAM29	10 kb *Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM29 in pBluescript II KS (+), Km^r, Ap^r	"
pAM37	6 kb *Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM37 in pBR325, Km^r, Tc^r, Ap^r, Cm^r	"
pAM52	5 kb *Eco*R I fragment carrying Tn*5* and flanking sequences of mutant AM52 in pBluescript II KS (+), Km^r, Ap^r	"
PLAFR3	IncP, Mob+, *LacZ*α, Tc^r, *cos*	Staskawicz *et al.*, 1987
PLAFR3XhoI	pLAFR3 with a *Xho*I site added to the *Bam*HI site using an adaptator	This study
pBC/Op4Δ	*Bam*HI-*PstI* fragment from pALB540 cloned between *Bam*HI and *Pst*I sites of pBCKS(+)	"
pBC/Op4ΔXhoI	PBC/Op4Δ with a *XhoI* site created by directed mutagenesis upstream from the *Bfr*I site	"
pBC/Op4Δ/XALB2	*Eco*RI DNA fragment from pAC389.1 cloned into the *Eco*RI site of pBC/Op4ΔXhoI	"
pBC/Op3-4/XALB2	*Bfr*I DNA fragment from pALB540 cloned into the *Bfr*I site of pBC/Op4Δ/XALB2	"
pBKS/XALB3	*Sal*I DNA fragment from pEV639 cloned into the *Sal*I site of pBluescript II KS (+)	"
pBKS/XALB3XhoI	pBKS/XALB3 with a *Xho*I site created by directed mutagenesis to substitute the *Sal*I site located on the *Kpn*I side of the polylinker	"
pBKS/Op3-4/XALB2-3	*Xho*I DNA fragment from pBC/Op3-4Δ/XALB2 cloned into the *Sal*I site of pBKS/XALB3XhoI	"
pOp3-4/XALB2-3	*Xho*I DNA fragment from pBKS/Op3-4/XALB2-3 cloned into the *Xho*I site of pLAFR3XhoI	"
pEValbXXII	*albXXII* in fusion with LacZ in pUFR043, Gm^r, Km^r	"
pEVHtpG	*E. coli htpG* in fusion with LacZ in pUFR043, Gm^r, Km^r	"
PGemT	ColE1 replicon, Ap^r, LacZα, single 3'-T overhangs at the insertion site	Promega
PGemT/albXXII	PCR fragment containing *albXXII* cloned into pGemT	This study
PGemT/albXXII bis	*Bgl*II-*Sal*IDNA fragment from pBKS/XALB3 cloned between the *BglII* and *Sal*I sites of pGemT/albXXII	"
PGemT/HtpG	PCR fragment containing the *E. coli htpG* gene cloned into pGemT	"

Table 1 : Bacterial strains and plasmids used in this study		
	Relevant characteristics[a]	**Reference or source**
DNA Fragment		
PR37	1.1 kb *Hind* III-*Hind* III from pAM37	"

[a] Ap^r, Cm^r, Gm^r, Km^r, Rif^r, Sp^r, Tc^r: resistant to ampicilin, chloramphenicol, gentamycin, kanamycin, rifampicin, spectinomycin, tetracycline, respectively. Tox-, deficient in albicidin production. Tn*5*-*gusA*, Tn*5*-*uidA1* Km^r Tc^r, forms transcriptional fusions. Alb^r, Ap^r, Gm^r, Rif^r and Tc^r: resistant to albicidin, ampicilin, gentamycin, rifampicin and tetracycline, respectively.

Table 2: Analysis of putative translational signals and location of all putative orfs identified in the XALB1 gene cluster

Intergenic spacing between consecutive ORFs in each putative operon	ORF	Potential RBS[a] (distance from start codon)	Start codon (position)	Stop codon (position)
Operon 1 (strand +)				
	albI	GAGGG (5 b)	TTG (30166)	TAG (50805)
45 b	*albII*	GAGGG (5 b)	ATG (50851)	TAA (51882)
ATG overlaps TAA	*albIII*	GAGGG (7 b)	ATG (51882)	TGA (52385)
GTG overlaps TGA	*albIV*	GAGG (7 b)	GTG (52382)	TAA (55207)
Operon 2 (strand -)				
	albV	GGAGG (8 b)	ATG (29929)	TAA (29210)
87 b	*albVI*	AAGG (4 b)	GTG (29122)	TGA (28262)
61 b	*albVII*	GAG (4 b)	ATG (28200)	TAG (25903)
7 b	*albVIII*	AGGTG (4 b)	ATG (25895)	TAA (24903)
20 b	*albIX*	GGTG (3 b)	ATG (24882)	TGA (19003)
Operon 3 (strand -)				
	albX	GGGGG (8 b)	ATG (14497)	TGA (14246)
81 b	*albXI*	AGGAAA (6 b)	ATG (14164)	TGA (13217)
5 b	*albXII*	GGCCTGA (5 b)	ATG (13211)	TAA (11856)
36 b	*albXIII*	GGGG (3 b)	ATG (11819)	TAA (10866)
12 b	*albXIV*	GGAG (8 b)	ATG (10853)	TAG (9363)
41 b	*albXV*	GGAA (6 b)	ATG (9321)	TAG (7567)
208 b	*albXVI*	GGAGG (4 b)	ATG (7358)	TAG (7092)
Operon 4 (strand +)				
	albXVII	GGGAGG (5 b)	TTG (14909)	TGA (17059)
274 b	*albXVIII*	GCTCAG (8 b)	ATG (17334)	TGA (17747)
Overlap (17 b)	*albXIX*	AGG (9 b)	ATG (17728)	TGA (18330)
41 b	*albXX*	GCAA (8 b)	ATG (18372)	TAG 18980)

[a]: Ribosomal Binding Site

Table 3: Deduced functions of the ORFs in the major albicidin biosynthetic cluster X-ALB1

ORF	Number of amino acids	Sequence homolog [a]	Proposed function[b, c]
Operon 1			
AlbI	6879	XabB (AAK15074)	Polyketide- peptide synthase
			PKS modules — PKS domains
			PKS-1 — AL ACP1
			PKS-2 — KS1 KR ACP2 ACP3
			PKS-3 — KS2 PCP1
			NRPS modules — NRPS domains
			NRPS-1 — C A PCP2
			NRPS-2 — C A PCP3
			NRPS-3 — C A PCP4
			NRPS-4 — C
Alibi	343	XabC (AAK15075)	*C*-methyltransferase
AlbIII	167	ComAB (CAA71583)	Activator of *alb* genes transcription
AlbIV	941	MycA (T44806) WbpG (E83253)	Peptide synthase
			NRPS module — NRPS domains
			NRPS-5 — A PCP5
Operon 2			
AlbV	239	Thp (AAK15074)	No function (transposition)
AlbVI	286	TcmP (AAA67510)	*O*-methyltransferase
AlbVII	765	HbaA (A58538)	4-hydroxybenzoate CoA ligase
albVIII	330	SyrP (AAB63253)	Regulation
AlbIX	1959	DhbF (CAB04779)	Peptide synthase
			NRPS modules — NRPS domains
			NRPS-6 — A PCP6
			NRPS-7 — C A PCP7
Operon 3			
AlbX	83	MbtH (O05821)	Unknown
AlbXI	315	SyrC (U25130)	Thioesterase
AlbXII	451	BoxB (AAK006000.1)	Unknown
albXIII	317	hp[d] (AAK25001)	Esterase
albXIV	496	ActII-2 (p46105)	Albicidin transporter
AlbXV	584	hp[d] (08390)	Carbamoyl transferase
AlbXVI	88	OrfA (AAC03166)	No function (transposition)
Operon 4			
albXVII	716	PabAB (CAC22117)	Para-amino benzoate synthase
Operon 5			
albXVIII	137	ADCL (AAG06352)	No function (not functional)
albXIX	200	McbG (P05530)	Immunity against albicidin
albXX	202	UbiC (S25660)	4-hydroxybenzoate synthetase

[a]Protein accession numbers in Genbank are given in parentheses.
[b]NRPS and PKS domains are abbreviated as follows: A, adenylation; ACP, acyl carrier protein; AL, acyl CoA ligase; C, condensation; KR, ketoreductase; KS, ketoacyl synthase; PCP, peptidyl carrier protein.
[c]Underlined domains are likely inactive due to the lack of highly conserved motifs.
[d]hypothetical protein

Table 4: Summary of results obtained from BLAST analyses.

Putative Alb protein	No. of aa residues	Protein homolog	Origin	Genbank accession #	Score	Expect	Identities	Positives	Gaps
AlbI	6879								
PKS-1		XabB (4801 aa) SafB (1770 aa)	*Xanthomonas albilineans* *Myxococcus xanthus*	AAK15074 AAC44128	1352 bits (3498) 231 bits (589)	0.0 2e-59	730/730 (100%) 175/532 (32%)	730/730 (100%) 269/532 (49%)	- 23/532 (4%)
PKS-2		XabB (4801 aa) PksM (4273 aa)	*X. albilineans* *Bacillus subtilis*	AAK15074 CAB13603	3464 bits (8983) 887 bits (2292)	0.0 0.0	1882/1882 (100%) 626/1896 (33%)	1882/1882 (100%) 938/1896 (49%)	- 140/1896=7%
PKS-3		XabB (4801 aa) PksM (4273 aa)	*X. albilineans* *B. subtilis*	AAK15074 CAB13603	1274 bits (3296) 577 bits (1486)	0.0 e-163	653/653 (100%) 293/584 (50%)	653/653 (100%) 391/584 (66%)	- 17/584 (2%)
NRPS-1		XabB (4801 aa) NosA (4379 aa)	*X. albilineans* *Nostoc* sp.	AAK15074 AF204805	1934 bits (5010) 618 bits (1594)	0.0 e-176	1035/1046 (99%) 398/1104 (36%)	1039/1046 (99%) 586/1104 (53%)	- 86/1104 (7%)
NRPS-2		NosA (4379 aa) Peptide synthase (5060 aa)	*Nostoc sp* *Anabaena* sp.	AF204805 CAC01604	416 bits (1069) 402 bits (1034)	e-115 e-111	337/1127 (29%) 315/1073 (29%)	496/1127 (43%) 479/1073 (44%)	128/1127 (11%) 114/1073 (10%)
NRPS-3		XabB (4801 aa) NosA (4379 aa)	*X. albilineans* *Nostoc* sp.	AAK15074 AF204805	1847 bits (4784) 610 bits (1573)	0.0 e-173	997/1044 (95%) 392/1069 (36%)	1007/1044 (96%) 571/1069 (52%)	- 86/1069 (8%)
NRPS-4		XabB (4801 aa) NosC (3317 aa)	*X. albilineans* *Nostoc sp.*	AAK15074 AAF17280	889 bits (2297) 240 bits (613)	0.0 2e-62	468/468 (100%) 156/438 (35%)	468/468 (100%) 229/438 (51%)	- 20/438 (4%)
AlbII	343	XabC (343 aa) MtmMII (326 aa) TcmO (339 aa)	*X. albilineans* *Streptomyces argillaceus* *S. glaucescens*	AAK15075 AAD55584 P39896	633 bits (1633) 144 bits (361) 81.7 bits (199)	0.0 1e-34 1e-14	343/343 (100%) 98/323 (30%) 79/314 (25%)	343/343 (100%) 154/323 (47%) 140/314 (44%)	- 4/323 (1%) 12/314 (3%)
AlbIII	167	*comA* operon protein 2 (136 aa) ComAB (116 aa)	*E. coli* *Bacillus licheniformis*	AAC74756 CAA71583	133 bits (335) 97.6 bits (242)	1e-30 8e-20	68/135 (50%) 53/111 (47%)	89/135 (65%) 68/111 (60%)	- 1/111 (0%)
AlbIV	941								
PKS-4		BA3 (6359 aa) WbpG (377 aa)	*B. licheniformis* *Pseudomonas aeruginosa*	AAC06348 E83253	361 bits (926) 81.6 bits (200)	2e-98 4e-15	190/441 (43%) 44/119 (36%)	267/441 (60%) 70/119 (57%)	14/441 (3%) 4/119 (3%)
AlbV	239	Thp (240 aa) IS transposase (260 aa)	*X. albilineans* *Yersinia pestis*	nd AAC82714	nd 160 bits (404)	0.0 1e-38	240/240 (100%) 87/183 (47%)	240/240 (100%) 122/183 (66%)	- 2/183 (1%)
AlbVI	286	Hypothetical protein TcmP (276 aa)	*Mycobacterium tuberculosis* *Pasteurella multocida*	AAK46042 AAK03406	138 bits (347) 36.6 bits (83)	6e-32 0.24	92/224 (41%) 32/132 (28%)	125/224 (55%) 65/132 (49%)	18/224 (8%) 29/197 (6%)

Table 4: Summary of results obtained from BLAST analyses.

Putative Alb protein	No. of aa residues	Protein homolog	Origin	Genbank accession #	Score	Expect	Identities	Positives	Gaps
AlbVII	765	4-hydroxybenzoate-CoA ligase (539 aa)	*Rhodopseudomonas palustris*	AAA62604	203 bits (513)	5e-51	156/492 (31%)	242/492 (48%)	31/492 (6%)
AlbVIII	330	SyrP Like (339 aa) SyrP (353 aa)	*S. verticillus* *Pseudomonas syringae*	AF210249 AAB63253	245 bits (619) 182 bits (458)	6e-64 5e-45	130/309 (42%) 106/306 (34%)	182/309 (58%) 155/306 (50%)	2/309 (0%) 4/306 (1%)
AlbIX	1959								
NRPS-6		XabB (4801 aa) DhbF (1278 aa)	*X. albilineans* *B. subtilis*	AAK15074 CAB15186	481 bits (1239) 354 bits (908)	e-135 1e-96	286/608 (47%) 222/608 (36%)	374/608 (61%) 341/608 (55%)	23/208 (3%) 21/608 (3%)
NRPS-7		XabB (4801 aa) NosA (4379 aa)	*X. albilineans* *Nostoc* sp.	AAK15074 AF204805	874 bits (2258) 551 bits (1420)	0.0 e-155	515/1110 (46%) 388/1148 (33%)	682/1110 (61%) 583/1148 (49%)	52/1110 (4%) 84/1148 (7%)
AlbX	83	Hypothetical protein (72 aa) MbtH (71 aa)	*P. aeruginosa* *M. tuberculosis*	AAG05800 CAB08480	75.6 bits (185) 59 bits (142)	1e-13 9e-09	34/61 (55%) 25/55 (45%)	44/61 (71%) 37/55 (66%)	- -
AlbXI	315	SyrC (433 aa) Hydrolase (261 aa)	*P. syringae* *S. coelicolor*	AAA85161 CAA16200	34.4 bits (78) 34 bits (77)	1.9 2.9	23/93 (24%) 19/60 (31%)	40/93 (42%) 30/60 (49%)	- -
AlbXII	451	BoxB (473 aa	*Azoarcus evansii*	AAK00599	293 bits (751)	3e-78	174/448 (38%)	243/448 (53%)	12/448 (2%)
AlbXIII	317	Hypothetical protein (335 aa) Plasma PAF acetylhydrolase (444 aa)	*Caulobacter crescentus* *Canis familiaris*	AAK25001 AAC48484	99.5 bits (247) 37.5 bits (86)	5e-200	88/296 (29%) 43/156	125/296 (41%) 56/156	5/296 (1%) 44/156 (28%)
AlbXIV	496	Putative trans-membrane efflux protein (505 aa) AlbF, putative albicidin efflux pump (496 aa)	*S. coelicolor* *X. albilineans*	CAB90983 AF403709	225 bits (574) 736 bits (1900)	0	154/465 (33%) 496/496 (100%)	240/465 (51%) 496/496 (100%)	8/465 (1%) -
AlbXV	584	Probable carbamoyl transferase (585 aa) BlmD (545 aa)	*P. aeruginosa* *S. verticillus*	AAG08390 AAG02370	201 bits (513) 192 bits (506)	1e-50 1e-47	158/458 (34%) 149/441 (33%)	222/458 (47%) 209/441 (46%)	39/458 (8%) 33/441 (7%)
AlbXVI	88	Transposase (363 aa) Transposase OrfA (88 aa)	*X. axonopodis* *Desulfovibrio vulgaris*	AF263433 AAC03166	64.8 bits (157) 61.0 bits (147)	2e-10 3e-09	27/45 (60%) 29/54 (53%)	40/45 (88%) 38/54 (69%)	- -

Table 4: Summary of results obtained from BLAST analyses.

Putative Alb protein	No. of aa residues	Protein homolog	Origin	Genbank accession #	Score	Expect	Identities	Positives	Gaps
AlbXVII	716	Para-aminobenzoate synthase (723 aa)	*Streptomyces griseus*	CAC22117	503 bits (1295)	e-141	302/699 (43%)	409/699 (58%)	36/699 (5%)
AlbXVIII	137	4-amino-4-deoxychorismate lyase (271 aa)	*P. aeruginosa*	AAG06352	81.4 bits (200)	4e-15	46/105 (43%)	65/105 (61%)	-
AlbXIX	200	McbG (187 aa)	*E. coli*	CAA30724	60.5 bits (145)	9e-09	36/141 (25%)	58/141 (40%)	5/141 (3%)
AlbXX	202	4-hydroxybenzoate synthase (202 aa)	*E. coli*	AAC77009	45.6 bits (107)	5e-04	42/161 (26%)	21/161 (13%)	-
AlbXXI	278	XabA (278aa)	*X. albilineans*	AAG28384	430 bits (1106)	0	278/278 (100%)	278/278 (100%)	-
AlbXXII	634	Heat shock protein HtpG (634)	*P. aeruginosa*	AAG04985	1051 bits (2688)	0	523/634 (82%)	588/634 (92%)	-
		Heat shock protein HtpG (624)	*E. coli*	AAC73575	743 bits (1899)	0	376/624 (60%)	476/624 (76%)	4/624 (0%)

Table 5 : Comparison of conserved sequences in C domains of peptide synthestases and in putative C domains of the Alb modules

Core	Sequences conserved in peptide synthetases*	Sequence	Alb module
C1	SxAQxR (L/M) (W/Y) xL	TYAQERLWLV STAQERMWFL *SYAQERLWLV* SLFQERLWFV SYQQERLWFV	NRPS-1 NRPS-2 NRPS-3 NRPS-4 NRPS-7
C2	*RHExLRTxF*	*RHEVLRTRF* RHAVLRTHF RHEILRTRF RHETLRTRI	NRPS-1 and NRPS-3 NRPS-2 NRPS-4 NRPS-7
C3	*MHHxISDG (W/V) S*	IHHIISDGWS IHHIVFDGWS MHHLIYDAWS MHHIICDGWS	NRPS-1 and NRPS-3 NRPS-2 NRPS-4 NRPS-7
C4	YxD (F/Y) AVW	YADYALW YADYARW YADYAIW YADYATW	NRPS-1 and NRPS-3 NRPS-2 NRPS-4 NRPS-7
C5	(I/V) GxFVNT (Q/L) (C/A) xR	*IGFFINILPLR* IGLFVNTLAVR IGFFVNILAVR	NRPS-1, NRPS-3 and NRPS-4 NRPS-2 NRPS-7
C6	(H/N) QD (Y/V) PFE	HQSVPFE HQDVPFE NQALPFE HRALPFE	NRPS-1 and NRPS-3 NRPS-2 NRPS-4 NRPS-7
C7	RDxSRNPL	RDSSQIPL RDTARNPL RDTSRIPL RDSSQIPL	NRPS-1 and NRPS-3 NRPS-2 NRPS-4 NRPS-7

*Sourced from Marahiel *et al.*, 1997

Table 6 : Comparison of conserved sequences in A domains of peptide synthestases and in putative A domains of the Alb modules			
Core	**Sequences conserved in peptide synthetases***	**Sequence**	**Alb module**
A1	L(T/S)YxEL	WSYAQL LSYAQL MSYGQL FSYRQL LSYAQL	NRPS-1 and NRPS-3 NRPS-2 NRPS-5 PKS-4 NRPS-6 and NRPS-7
A2	LKAGxAYL(V/L)P(L/I)D	FKAGACYVPID SLCGAASVLID MKAGAAYVPID LAGGLVFAPIN LKAGGCYVPLD	NRPS-1 and NRPS-3 NRPS-2 NRPS-5 PKS-4 NRPS-6 and NRPS-7
A3	LAYxxYTSG(S/T)TGxPKG	LACVMVTSGSTGRPKG ?TRTIMVESGSLSSRLL? PVYCIYTSGSTGSPKG PAVMICTSGSTGTPKA *LAYVMYTSGSTGRPKG*	NRPS-1 and NRPS-3 NRPS-2 NRPS-5 PKS-4 NRPS-6 et NRPS-7
A4	FDxS	FAVS FDAA FDLT FAYG FAIS	NRPS-1 and NRPS-3 NRPS-2 NRPS-5 PKS-4 NRPS-6 and NRPS-7
A5	NxYGPTE	NNYGCTE ?AAYGNAE? NEYGPTE DGIGCTE YIYGCTE	NRPS-1 and NRPS-3 NRPS-2 NRPS-5 PKS-4 NRPS-6 and NRPS-7
A6	GELxIxGxG(V/L)ARGYL	GELHVHSVGMARGYW np GQIHIGGAGVAIGYV GSLWVRGNTLTRGYV GEVHIESLGITHGYW	NRPS-1 and NRPS-3 NRPS-2 NRPS-5 PKS-4 NRPS-6 and NRPS-7
A7	Y(R/K)TGDL	*YKTGDM* ?YKTDAL? YASGDL ?FDTRDL? YRTGDM	NRPS-1 and NRPS-3 NRPS-2 NRPS-5 PKS-4 NRPS-6 and NRPS-7
A8	GRxDxQVKIRGxRIELGEIE	GRQDFEVKVRGHRVDTRQVE ?GSLDVQSRIDDPRIDLCVVE? GRKDSQIKLRGYRIELGEIE ?GRMGSAIKINGCWLSPETLE? GRRDYEVKVRGYRVDVRQVE	NRPS-1 and NRPS-3 NRPS-2 NRPS-5 PKS-4 NRPS-6 and NRPS-7
A9	LPxYM(I/V)P	LPTYMLP ?LPDYLLP? LPEYMLP ?LGKHHYP? LPTYMLP	NRPS-1 and NRPS-3 NRPS-2 NRPS-5 PKS-4 NRPS-6 and NRPS-7

Table 6 : Comparison of conserved sequences in A domains of peptide synthestases and in putative A domains of the Alb modules			
Core	**Sequences conserved in peptide synthetases***	**Sequence**	**Alb module**
A10	NGK(V/L)DR	*NGKLDR* ?HGRVDL? NGKVNR ?SGKVIR? NGKLDT	NRPS-1 and NRPS-3 NRPS-2 NRPS-5 PKS-4 NRPS-6 and NRPS-7

*Sourced from Marahiel *et al.*, 1997
?: non conserved sequences
np: not present

Table 7: Comparison of conserved sequences in PCP and TE domains of peptide synthestases and in putative PCP and TE domains of the Alb modules			
Domain	**Sequences conserved in peptide synthetases***	**Sequence**	**Alb module (domain)**
PCP	DxFFxxLGG(H/D)S(L/I)	D-FFAVGGHSVL DNFFALGGHSL*S* *DNFFELGGHSVL* DNFFELGGHSLS DNFFNLGGHSLL	PKS-3 (PCP1) NRPS-1 and NRPS-3 (PCP2 and PCP4) NRPS-2 (PCP3) NRPS-5 (PCP5) NRPS-6 and NRPS-7 (PCP6 and PCP7)
TE	G(H/Y)SxG	GWSSG	NRPS-7

*Sourced from Marahiel *et al.*, 1997

Table 8.

Domains	Position in GsrA (Phe) and variability									
	235	236	239	278	299	301	322	330	331	517
	0	+/-	++	++	++	+/-	++	+/-	++	0
Alb NRPS-1	**A**	**V**	K	**Y**	**V**	**A**	**N**	D	A	**K**
Alb NRPS-3	**A**	**V**	K	**Y**	**V**	**A**	**N**	D	A	**K**
TyrB-M1 (Pro)	D	**V**	Q	S	**I**	**A**	**N**	V	V	**K**
VirS (Pro)	D	**V**	Q	**Y**	A	**A**	H	V	M	**K**
HVCL	**G**	A	L	H	**V**	V	G	S	I	**K**
Alb NRPS-6	A	**I**	K	**Y**	**F**	**S**	**I**	D	**M**	**K**
Alb NRPS-7	A	**I**	K	**Y**	**F**	**S**	**I**	D	**M**	**K**
VirS (Pro)	D	**V**	Q	**Y**	A	A	H	V	**M**	**K**
EntF-M1 (Ser)	D	**V**	W	H	**F**	**S**	**L**	V	D	**K**
β-Ala code	V	D	W	V	I	**S**	**L**	A	D	**K**
Alb NRPS-5	D	L	T	K	I	G	E	V	G	K
BacC-M5 (Asn)	D	L	T	K	I	G	E	V	G	K
TyrC-M1 (Asn)	D	L	T	K	I	G	E	V	G	K
Asn code	D	L	T	K	L	G	E	V	G	K

Table 9: Complementation studies of Xa23RI insertion mutants

Donor	Recipient				
	AM12	AM13	AM36	AM10	AM15
pEV639	+	+	+	-	-
pEValbXXII	+	+	+	-	-
pEVHtpG	+	+	+	-	-
pALB639	+	+	+	-	-
pUFR043	-	-	-	-	-
none	-	-	-	-	-

+ : restoration of albicidin production by alb⁻ mutant, - : no complementation. All experiments were performed at least in duplicate with at least 2 exconjugants obtained from two independent triparental conjugations.

Table 10: Albicidin production assays with *X. axonopodis* pv. vesicatoria exconjugants harbouring different plasmids: analysis of growth inhibition of *E. coli* DH5αKT (susceptible to albicidin) and DH5αAlb^{r}KT (resistant to albicidin) in assays performed with different antibiotic combinations (no antibiotic, tetracycline only, kanamycin only and tetracycline+kanamycin).

Bioassay medium containing	Tester strain	Combination of plasmids: pUFR043 and pLAFR3	pUFR043 and pOp3-4/XALB2-3	PALB571 and pLAFR3	PALB571 and pOp3-4/XALB2-3
No antibiotic	DH5αKT	-	-	-	-
	DH5αAlb^{r}KT	-	-	-	-
Tetracycline	DH5αKT	-	-	-	+
	DH5αAlb^{r}KT	-	-	-	-
Kanamycin	DH5αKT	+	+	+	+
	DH5αAlb^{r}KT	+	+	+	+
Tetracycline+kanamycin	DH5αKT	+	+	+	+
	DH5αAlb^{r}KT	+	+	+	+

+: presence of a growth inhibition zone

All experiments were performed at least in duplicate with at least 2 exconjugants obtained from two independent triparental conjugations.

REFERENCES

1. Albertini, A.M., Caramori, T., Scoffone, F., Scotti, C. & Galizzi, A. (1995). Sequence around the 159 degree region of the *Bacillus subtilis* genome: the pksX locus spans 33.6 kb. *Microbiology* **141**, 299-309.

2. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein data base search programs," *Nucleic Acids Res.* **25**, 3389-3402.

3. Aparicio, J.F., Molnar, I., Schwecke, T., Konig, A., Haydock, S.F., Khaw, L.E., Staunton, J. & Leadlay, P.F. (1996). Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169**, 9-16.

4. Astua-Monge, G., Minsavage, G.V., Stall, R.E., Davis, M.J., Bonas, U. and Jones, J.B. (2000). Resistance of tomato and pepper to T3 strains of *Xanthomonas campestris* pv. vesicatoria is specified by a plant-inducible avirulence gene. Mol. Plant Microbe Interact. 13:911-21.

5. August, P.R., Tang, L., Yoon, Y.J., Ning, S., Muller, R., Yu, T.W., Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X., Hutchinson, C.R. & Floss, H.G. (1998). Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. *Chem. Biol.* **5**, 69-79.

6. Bardwell, J.C. & Craig, E.A. (1988). Ancient heat shock gene is dispensable. *J. Bacteriol.* **170**, 2977-83.

7. Birch, R.G. (2001). *Xanthomonas albilineans* and the antipathogenesis approach to disease control. *Mol. Plant Pathol.* **2**, 1-11.

8. Birch, R.G. & Patil, S.S. (1983). The relation of blocked chloroplast differentiation to sugarcane leaf scal disease. *Phytopathology* **73**, 1368-1374.

9. Birch, R.G. & Patil, S.S. (1985a). June 1985. Antibiotic and process for the production thereof. U. S. Patent No. 4,525,354.

10. Birch, R.G. & Patil, S.S. (1985b) Preliminary characterization of an antibiotic produced by *Xanthomonas albilineans* which inhibits DNA synthesis in *Escherichia coli*. *J. Gen. Microbiol.* **131**, 1069-1075.

11. Birch, R.G. & Patil, S.S. (1987a). Correlation between albicidin production and chlorosis induction by *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. *Physiol. Mol. Plant Pathol.* **30**, 199-206.

12. Birch, R.G. & Patil, S.S. (1987b). Evidence that an albicidin-like phytotoxin induces chlorosis in sugarcane leaf scald disease by blocking plastid DNA replication. *Mol. Plant Pathol.* **30**, 207-214.

13. Borkovich, K.A., Farrelly, F.W., Finkelstein, D.B., Taulien, J. & Lindquist, S. (1989). Hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol. Cell. Biol.* **9**, 3919-30.

14. Brendel, V. & Trifonov, E. N. (1984). A computer algorithm for testing potential prokaryotic terminators. *Nucl. Acids Res.* **12** 4411-4427.

15. Cane, D.E. (1997). A special thematic issue on polyketide and nonribosomal polypeptide biosynthesis. *Chem. Rev.* **97**, 2463-2706.

16. Cane, D.E., Walsh, C.T. & Khosla, C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**, 63-8.

17. Cane, D.E. & Walsh, C.T. (1999). The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chem. Biol.* **6**, R319-R325.

18. Challis, G.L., Ravel, J. & Townsend, C.A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peotide synthetase adenylation domains. *Chem. Biol.* **7**, 211-224.

19. Conti, E., Stachelhaus, T. Marahiel, M.A. & Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174-4183.

20. Criado, L.M. , Martin, J.F. & Gil, J.A. (1993). The pab gene of *Streptomyces griseus*, encoding p-aminobenzoic acid synthase, is located between genes possibly involved in candicidin biosynthesis. *Gene* **126**, 135-139.

21. Cutforth, T. & Rubin, G.M. (1994). Mutations in Hsp83 and cdc37 impair signaling by the sevenless receptor tyrosine kinase in Drosophila. *Cell.* **77**, 1027-36.

22. De Feyter, R. & Gabriel, D.W. (1991). Use of cloned DNA methylase genes to increase the frequency of transfer of foreign genes into *Xanthomonas campestris pv. malvacearum. J. Bacteriol.* 173, 336-342.

23. von Döhren, H., Dieckmann, R. & Pavela-Vrancic, M. (1999). The nonribosomal code. *Chem. Biol.* **6**, R273-R279.

24. Du, L., Sanchez, C., Chen, M., Edwards, D.J. & Shen, B. (2000). The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonrobosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* 7, 623-642.

25. Duitman, E.H., Hamoen, L.W., Rembold, M., Venema, G., Seitz, H., Saenger, W., Bernhard, F., Reinhardt, R., Schmidt, M., Ullrich, C., Stein, T., Leenders, F. & Vater, J. (1999). The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc. Natl. Acad. Sci. USA* **96**, 13294-13299.

26. Gabriel, D.W & De Feyter, R. (1992). RFLP analyses and gene tagging for bacterial identification and taxonomy. In *Molecular Plant Pathology: A Practical Approach.* (Gurr, S.J., McPherson, M.J. & Bowles, D.J. Ed.), vol 1, pp51-66, IRL Press, Oxford.

27. Garrido, M.C., Herrero, M., Kolter, R. & Moreno, F. (1988). The export of the DNA replication inhibitor Microcin B17 provides immunity for the host cell. *EMBO J.* **7**, 1853-1862.

28. Gehring, A.M., DeMoll, E., Fetherston, J.D., Mori, I., Mayhew, G.F., Blattner, F.R., Walsh, C.T. & Perry, R.D. (1998). Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem. Biol.* **5**, 573-86.

29. Guenzi, E., Galli, G., Grgurina, I., Gross, D.C. & Grandi, G. (1998). Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* **273**, 32857-32863.

30. Huang, G., Zhang, L. & Birch, R.G. (2000a). Analysis of the genes flanking *xabB*: a methyltransferase gene is involved in albicidin biosynthesis in *Xanthomonas albilineans*. *Gene* **255**, 327-333.

31. Huang, G., Zhang, L. & Birch, R.G. (2000b). Albicidin antibiotic and phytotoxin biosynthesis in *Xanthomonas albilineans* requires a phosphopantetheinyl transferase gene. *Gene* **258**, 193-199.

32. Huang, G., Zhang, L. & Birch, R.G. (2001). A multifunctional polyketide-peptide synthetase essential for albicidin biosynthesis in *Xanthomonas albilineans*. *Microbiology* **147**, 631-642.

33. Jakob, U., Lilie, H., Meyer,I. & Buchner, J. (1995). Transient interaction of Hsp90 with early unfolding intermediates of citrate synthase. Implications for heat shock *in vivo*. *J. Biol. Chem.* **270**, 7288-94.

34. Konz, D., Klens, A., Schorgendorfer, K. & Marahiel, M.A. (1997). The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC10716: molecular characterization of three multi-modular peptide synthetase. *Chem. Biol.* **4**, 927-937.

35. Leong, S.A., Ditta, G.S. & Helinski, D.R. (1982). Heme biosynthesis in *Rhizobium*: identification of a cloned gene coding for aminolevulinic acid synthetase from *Rhizobium meliloti*. *J. Biol. Chem.* **257**, 8724-8730.

36. Li, Y.M., Milne, J.C., Madison, L.L., Kolter, R. & Walsh, C.T. (1996). From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. *Science*. **274**, 1188-93.

37. Liu, J., Duncan, K. & Walsh, C.T. (1989). Nucleotide sequence of a cluster of *Escherichia coli* enterobactin biosynthesis genes: identification of ent A and purification of its product 2,3-dihydro-2,3-dihydroxybenzoate deshydrogenase. *J. Bacteriol.* **171**, 791-798.

38. Marahiel, M.A. Stachelhaus, T. & Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**, 2651-2673.

39. Mayorga, M.E. & Timberlake, W.E. (1992). The developmentally regulated *Aspergillus nidulans wA* gene encodes a polypeptide homologous to polyketide and fatty acid synthases. *Mol. Gen. Genet.* **235**, 205-212.

40. McCarthy, J.E. & Gualerzi, C. (1990). Translational control of prokaryotic gene expression. *Trends Genet.* **6**,78-85.

41. McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M. & Ashley, G. (1999). Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. USA* **96**, 1846-1851.

42. Milne, J.C., Roy, R.S., Eliot, A.C., Kelleher, N.L., Wokhlu, A., Nickels, B. & Walsh, C.T. (1999). Cofactor requirements and reconstitution of microcin B17 synthetase: a multienzyme complex that catalyzes the formation of oxazoles and thiazoles in the antibiotic microcin B17. *Biochemistry.* **38**, 4768-81.

43. Mohamed, I.S., Rott, P., Davis, M.J. & Chatenet, M. (1996). Differentiation of *Xanthomonas albilineans* strains based on multiplication of the pathogen in sugarcane varieties. In *Proceedings XXII Congress of the International Society of Sugarcane Technologists*. (Cock J.H. and Brekelbaum T. Ed.), vol. 2, pp486-492, Cartagena De Indias, Colombia, 5-14 September 1995.

44. Mohamed, M.E., Zaa, A., Ebenau-Jehle, C. & Fuchs, G. (2001). Reinvestigation of a new type of aerobic benzoate metabolism in the proteobacterium *Azoarcus evansii*. *J. Bacteriol.* **183**, 1899-1908.

45. Motamedi, H. & Shafiee, A. (1998). The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *Eur. J. Biochem.* **256**, 528-534.

46. Normark, S., Bergström, S., Edlund, T., Grundström, T., Jaurin, G., Lindberg, F.P. & Olsson O. (1983). Overlapping genes. *Annu. Rev. Genet.* **17**, 499-525.

47. Paitan, Y., Orr, E., Ron, E.Z. & Rosenberg, E. (1999). Genetic and functional analysis of genes required for the post-modification of the polyketide antibiotic TA of *Myxococcus xanthus*. *Microbiology* **145**, 3059-67.

48. Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E. & Khosla C. (2001). Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790-1792.

49. Quadri, L.E., Sello, J., Keating, T.A., Weinreb, P.H. & Walsh, C.T. (1998). Identification of a *Mycobaterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem. Biol.* **5**, 631-645.

50. Ricaud, C. & Ryan, C.C. (1989). Leaf scald. In *Diseases of sugarcane: major diseases*. (Ricaud, C. Egan, B.T., Gillaspie, Jr., A.G. & Hughes, C.G. ed.), pp. 39-58, Elsevier Science Publishers B.V., Amsterdam.

51. Rodriguez, E. & McDaniel, R. (2001). Combinatorial biosynthesis of antimicrobials and other natural products. *Curr. Opin. Microbiol.* **4**, 526-34.

52. Rott, P.C., Costet, L., Davis, M.J., Frutos, R. & Gabriel D.W. (1996). At least two separate gene clusters are involved in albicidin production by *Xanthomonas albilineans*. *J. Bacteriol.* **178**, 4590-4596.

53. Rott, R. & Davis, M.J. (2000). Leaf scald. In *A guide to sugarcane diseases*. (Rott, P., Bailey, R.A., Comstock, J.C., Croft, B.J. and Saumtally, A.S. ed.), pp. 38, Cirad/Issct, Montpellier.

54. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular cloning: A laboratory Manual* (2nd edn), Cold Spring Harbor Lab. Press, Plainview, NewYork.

55. Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. USA* **74**, 5463-5467

56. Siebert, M., Bechthold, A., Melzer, M., May, U., Berger, U. Schroder, G., Schroder, J., Severin, K. & Heide, L. (1992). Ubiquinone biosynthesis. Cloning of the genes coding for chorismate pyruvate-lyase and 4-hydrobenzoate octaprenyl transferase from *Escherichia coli*. *FEBS Lett.* **307**, 347-350.

57. Stachelhaus, T., Mootz, H.D. & Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493-505.

58. Staskawicz, B. Dahlbeck, D., Keen, N. and Napoli, C. (1987). Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. glycenea. J. Bacteriol. 169: 5789-5794.

59. Takano, Y., Kubo, Y., Shimizu, K., Mise, K., Okuno, T. & Furusawa, I. (1995). Structural analysis of PKS1, a polyketide synthase gene involved in melanin biosynthesis in *Colletotrichum lagenarium*. *Mol. Gen. Genet.* **249**, 162-167.

60. Tang, L., Yoon, Y.J., Choi, C.Y. & Hutchinson, C.R. (1998). Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis by *Amycolatopsis mediterranei*. *Gene* **216**, 255-65.

61. Thomas, J.G. & Baneyx, F. (2000). ClpB and HtpG facilitate *de novo* protein folding in stressed *Escherichia coli* cells. *Mol. Microbiol.* **36**, 1360-70.

62. Valdez, F., Gonzalez-Ceron, G., Kieser, H.M., Servin-Gonzalez, L. (1999). The *Streptomyces coelicolor* A3(2) lipAR operon encodes an extracellular lipase and a new type of transcriptional regulator. *Microbiology* **145**, 2365-74.

63. Wall, M.K. & Birch, R.G. (1997). Genes for albicidin biosynthesis and resistance span at least 69 kb in the genome of *Xanthomonas albilineans*. *Lett. Appl. Microbiol.* **24**, 256-260.

64. Yakimov, M.M., Kroger, A., Slepak, T.N., Giuliano, L., Timmis, K.N. & Golyshin, P.N. (1998). A putative lichenysin A synthase operon in *Bacillus licheniformis*: initial characterization. *Biochim. Biophys. Acta.* **1399**, 141-153.

65. Yazgan, A., Ozcengiz, G. & Marahiel M.A. (2001). Tn*10* insertional mutations of *Bacillus subtilis* that block the biosynthesis of bacilysin. *Biochim. Biophys. Acta.* **1518**, 87-94.

66. Yu, J.H. & Leonard, T.J. (1995). Sterigmatocystin biosynthesis in *Aspergillus nidulans* requires a novel type I polyketide synthase. *J. Bacteriol.* **177**, 4792-4800.

67. Yu, S., Fiss, E., & Jacobs, W.R. (1998). Analysis of the exochelin locus in *Mycobacterium smegmatis*: biosynthesis genes have homology with genes of the peptide synthetase family. *J. Bacteriol.* **180**, 4676-4685.

68. Zhang, L. & Birch, R.G. (1997). The gene for albicidin detoxification from *Pantoea dispersa* encodes an esterase and attenuates pathogenicity of *Xanthomonas albilineans* to sugarcane. *Proc. Natl. Acad. Sci. USA* **94**, 9984-9989.

69. Zhang, J. H., Quigley, N.B. & Gross, D.C. (1995). Analysis of the syrB and syrC genes of Pseudomonas syringae pv. syringae indicates that

syringomycin is synthesized by a thiotemplate mechanism. *J. Bacteriol.* **177**, 4009-4020.

70. Zhang, J. H., Quigley, N. B. & Gross, D.C. (1997). Analysis of the syrP gene, which regulates syringomycin synthesis by *Pseudomonas syringae pv. syringae*. *Appl. Environ. Microbiol.* **63**, 2771-2778.

71. Zhang, L., Xu, J. & Birch, R.G. (1998). Factors affecting biosynthesis by *Xanthomonas albilineans* of albicidins antibiotics and phytotoxins. *J. Appl. Microbiol.* **85**, 1023-1028.

72. Zhang L., Xu, J. & Birch, R.G. (1999). Engineered detoxification confers resistance against a pathogenic bacterium. *Nat Biotechnol.* **17**:1021-1024.